

Education

Index in Dec
APR 6 1950

SCHOOL OF EDUCATION
LIBRARY

Review of Educational Research

VOL. XX, No. 1

FEBRUARY 1950

EDUCATIONAL AND PSYCHOLOGICAL TESTING

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

A Department of the

NATIONAL EDUCATION ASSOCIATION OF THE UNITED STATES

1201 Sixteenth St., N.W., Washington 6, D. C.

L
11
R454

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

THIS ASSOCIATION is composed of persons engaged in technical research in education, including directors of research in school systems, instructors in educational institutions, and research workers connected with private educational agencies.

Officers, February 1949—February 1950

President: HELEN M. WALKER, Professor of Education, Teachers College, Columbia University, New York 27, New York.

Vicepresident: ARTHUR E. TRAXLER, Associate Director, Educational Records Bureau, New York 19, New York.

Secretary-Treasurer: FRANK W. HUBBARD, Director, Research Division, National Education Association, Washington 6, D. C.

Executive Committee

Consists of five members: The president, vicepresident, secretary-treasurer, chairman of the Editorial Board, and the following immediate past president: WILLARD C. OLSON, Director of Research in Child Development in the University Elementary School, and Professor of Education, School of Education, University of Michigan, Ann Arbor, Michigan.

Editorial Board

Chairman and Editor:* HARRY A. GREENE, Professor of Education, State University of Iowa, Iowa City, Iowa.

ARNOLD E. JOYAL, President, Fresno State College, Fresno 4, California.

JOHN DALE RUSSELL, Director, Division of Higher Education, U. S. Office of Education, Washington 25, D. C.

The president; and the secretary-treasurer.

***Assistant Editor:** ALBERT N. HIERONYMUS, State University of Iowa, Iowa City, Iowa.

Applications for membership should be sent to the secretary-treasurer. Upon approval by the Executive Committee persons applying will be invited to become members.

Subscriptions to the REVIEW should be sent to the secretary-treasurer (note address above).

Orders for one or more publications, accompanied by funds in payment, should be sent to the American Educational Research Association, 1201 Sixteenth St., N. W., Washington 6, D. C. For a list of titles see the back inside cover page.

Active members of the Association pay dues of \$7 annually. Associate members pay \$6 annually. Of this amount \$5 is for subscription to the REVIEW. The REVIEW is published in February, April, June, October, and December. Beginning with the February 1949 issue single copies are priced at \$1.50.

Entered as second-class matter April 10, 1931, at the post office at Washington, D. C., under the Act of August 24, 1912.

Educ.

REVIEW OF EDUCATIONAL RESEARCH

*Official Publication of the American Educational Research Association.
Contents are listed in the Education Index.*

Copyright, 1950

By National Education Association of the United States, Washington, D. C.

Vol. XX, No. 1

February 1950

Educational and Psychological Testing

Reviews the literature for the three-year period, August 1946 thru July 1949. Volume XVII, Number 1, February 1947, covers the previous three-year period.

TABLE OF CONTENTS

Chapter	Page
Foreword	4
I. Overview of Educational and Psychological Testing, 1946 to 1949	5
DAVID SEGEL, <i>U. S. Office of Education, Washington, D. C.</i>	
J. RAYMOND GERBERICH, <i>University of Connecticut, Storrs</i>	
II. Construction and Educational Significance of Intelligence Tests	17
ETHEL L. CORNELL, <i>New York State Department of Education, Albany</i>	
ANNETTE GILLETTE, <i>Hartford Public Schools, Hartford, Connecticut</i>	
III. Construction and Educational Significance of Aptitude Tests ..	27
DEWEY B. STUIT, <i>State University of Iowa, Iowa City</i>	
IV. Construction and Educational Significance of Structured Inventories in Personality Measurement	38
ARTHUR E. TRAXLER, <i>Educational Records Bureau, New York</i>	
ROBERT JACOBS, <i>Educational Records Bureau, New York</i>	
V. Development and Educational Significance of Projective Technics in Personality Measurement	51
PERCIVAL M. SYMONDS, <i>Teachers College, Columbia University, New York</i>	
MARTHA G. HESSEL, <i>Teachers College, Columbia University, New York</i>	
VI. Measurement of Educational Achievement in the Schools	63
WARREN G. FINDLEY, <i>Educational Testing Service, Princeton, New Jersey</i>	
ALLAN B. SMITH, <i>University of Connecticut, Storrs</i>	

<i>Chapter</i>	<i>Page</i>
VII. Educational Achievement Measures in Scholarship and Award Contests	76
IRVING LORGE, <i>Teachers College, Columbia University, New York</i>	
ROSE KUSHNER, <i>College of the City of New York, New York</i>	
VIII. Measurement of Educational Achievement in Nonschool Educational Agencies and in Industry	81
J. RAYMOND GERBERICH, <i>University of Connecticut, Storrs</i>	
JAMES M. BURKE, <i>University of Connecticut, Storrs</i>	
IX. Construction and Validation of Educational Tests	87
ROBERT L. EBEL, <i>State University of Iowa, Iowa City</i>	
Index	99

This issue of the REVIEW was prepared by the
Committee on Educational and Psychological Tests

J. RAYMOND GERBERICH, *Chairman*, University of Connecticut, Storrs
ETHEL L. CORNELL, New York State Education Department, Albany
WARREN G. FINDLEY, Educational Testing Service, Princeton, New Jersey
IRVING LORGE, Teachers College, Columbia University, New York
DEWEY B. STUIT, State University of Iowa, Iowa City
ARTHUR E. TRAXLER, Educational Records Bureau, New York

with the assistance of

JAMES M. BURKE, University of Connecticut, Storrs
ROBERT L. EBEL, State University of Iowa, Iowa City
ANNETTE L. GILLETTE, Hartford Public Schools, Hartford, Connecticut
MARTHA G. HESSEL, Teachers College, Columbia University, New York
ROBERT JACOBS, Educational Records Bureau, New York
ROSE KUSHNER, College of the City of New York, New York
DAVID SEGEL, U. S. Office of Education, Washington, D. C.
ALLAN B. SMITH, University of Connecticut, Storrs
PERCIVAL M. SYMONDS, Teachers College, Columbia University, New York

FOREWORD

THIS issue of the REVIEW is the first of what probably will become a new topic devoting attention jointly to educational and psychological testing. It reestablishes the topic of the February 1933, December 1935, and December 1938 issues on "Educational Tests and Their Uses," in combination with the topic, "Psychological Tests and Their Uses," to which the October 1932, the June 1932, 1935, and 1938, and the February 1941, 1944, and 1947 numbers were devoted. Altho a chapter or two of the issues on methods of research and appraisal have typically dealt with tests, measurements, and related problems, and occasional chapters of the special subjectmatter numbers have been concerned with educational testing, the resumption of greater attention to educational measurement appears to be justified by recent developments in the field.

Testing in the armed forces receives only minor attention in this issue, inasmuch as the special December 1948 number dealt extensively with the highly significant developments in measurement resulting directly from World War II. On the other hand, attention is directed in this issue to subjects which have almost certainly not received integrated treatments in previous issues—achievement measurement conducted in, or at least with the cooperation of, the schools by certain nonschool agencies, both educational and commercial, and achievement and proficiency measurement carried on by industry. It was found desirable to cover considerably more than a three-year span in the treatment of scholarship and award contests in order to review the origins as well as the recent developments.

The organization of this issue more closely follows that of the previous issues titled "Psychological Tests and Their Uses" than it does the issues of the 1930's titled "Educational Tests and Their Uses." The field of educational and psychological testing is divided into three major areas: (a) general intelligence and aptitude measurement, (b) personality measurement, both by use of structured inventories and projective technics, and (c) achievement measurement, ranging from that conducted in and by the schools to that conducted in and by nonschool and noneducational agencies.

Chapter authors have dealt with the theoretical and developmental aspects as well as the applicational aspects of research in their respective areas, altho such divisions do not appear as discrete chapter sections. Several of the authors have also given direct consideration to recent trends as the measurement movement enters its second half-century of development.

Appreciation of the chairman is tendered to the committee members and other chapter authors for their excellent cooperation and their valuable contributions to this issue. Special appreciation goes to Warren G. Findley, who stepped in at a late date as senior author of a chapter when one of the chapter authors found it necessary to withdraw.

J. RAYMOND GERBERICH, *Chairman,*
Committee on Educational and Psychological Testing

CHAPTER I

Overview of Educational and Psychological Testing, 1946 to 1949

DAVID SEGEL and J. RAYMOND GERBERICH

THE three-year period covered by this issue of the REVIEW is of special significance in the history of educational measurements. It marks the approximate beginning of the second half-century of development in the objective measurement of human behavior. Originating in 1897 with the work of Rice, and continuing by devious routes, seemingly hampered by but probably benefiting from various obstacles in its path, the measurement movement recently entered its sixth decade of development. The purpose of this introductory chapter is to chart briefly the progress of this movement during the past fifty years and in somewhat greater detail to note current developments and trends.

Scates (43) traced the movement thru its first five decades: (a) the incubation period, 1897 to 1906, following the proposals of Rice and culminating in the publication of the first Binet scale; (b) the second period, 1907 to 1916, marked by the work of Thorndike and his students, the publication of the first standardized scales and achievement tests, and the fight for objectivity in testing educational achievement; (c) the third period, 1917 to 1926, one of rapid expansion for educational measurements, during which the exigencies of World War I contributed directly to the introduction of group intelligence tests and stimulated their postwar development, and during which the technics of measurement were closely scrutinized; (d) the fourth period, 1927 to 1936, characterized by direct attention to the objectives of instruction, to the evaluation of instructional outcomes as evidenced in human behavior, and by the issuance of many new and improved tests and testing technics for personality measurement as well as for intelligence and achievement measurement; and (e) the period closing the first half-century, 1937 to 1946, marked primarily by the developments resulting from World War II and the development of factor analysis technics.

Books and Monographs on Measurement

Altho the years 1941 to 1945 were characterized by highly significant developments in educational and psychological measurement, the postwar years of 1946 to 1949 have been marked by the issuance of many more books in this field than appeared during the two preceding three-year periods. Some of the books reported results of test development and related research in the armed forces and others were based on their authors'

experiences in various aspects of the war effort. Still others emanated from civilian sources, for not all of the measurement developments of the war period were connected with war needs.

General Publications

A supplement to her earlier bibliography of mental tests and rating scales was issued by Hildreth (28) in 1946. Buros issued his *Third Mental Measurements Yearbook* (8), which is actually the sixth publication in his series of yearbooks and the preceding test bibliographies. This 1047-page volume presented 713 original reviews written by 320 reviewers and excerpts from 66 reviews previously published elsewhere. A total of 663 tests were reviewed. In furtherance of his plan to have each commercially available test evaluated at least three times in the series of yearbooks, 30 percent of the tests were reviewed by two or more persons and a small percentage of tests were reviewed three, four, or even more than four times. The momentous nature of the Yearbook is further shown by the fact that 3368 references on the construction, validity, use, and limitations of tests were included. Buros also listed 549 books on measurement and closely related subjects and included excerpts from 785 reviews of them in 135 journals, to make available still another type of significant information for the measurement technician.

Intelligence and Aptitude Testing

In the field of intelligence testing, Mursell (35) issued a revision of his 1947 book and Goodenough (25) brought out a treatment which gave significant attention to the historical development as well as the modern principles and applications of mental testing. Cronbach (16) devoted attention to mental testing and also treated measurement of achievement, attitudes, interests, and various observational and projective technics. A second edition of Burt's book (10) on mental and scholastic tests was published.

Predictive values of testing received the attention of Froehlich and Benson (23) in public-school guidance programs, of Crawford and Burnham (15) at the college level, and of Stuit and others (48) for professional schools. Super (50) analyzed selected tests and inventories of widely varied types for their values in predicting vocational success. Attacking the same problem in terms of placement rather than of guidance and dealing with tests constructed to meet particular situations rather than with standardized instruments, Thorndike (51) drew widely upon experience in the Aviation Psychology Program for his quite technical book on personnel selection, while Stephenson (47) wrote on the selection of school pupils in an essay making wide use of the principles of educational and social psychology.

Educational Testing

Adkins (1) produced a guide for constructing paper-and-pencil and performance tests and for test validation technics. A manual of achievement test construction technics was issued by Weitzman and McNamara (55), and Burt (9) rewrote his handbook of tests. Krakower (32) applied tests and measurements to nursing education, while Anderson and Lindquist (3) issued a revision of their bulletin presenting test items in American history. Useful tools for measurement workers are the manual for adapting tests to machine scoring (30) and the exposition by Pease (37) on the use of various makes of machines for a variety of statistical computations.

An extensive exposition on the measurement of understanding was issued by Brownell and his colleagues (7), whose contribution stressed the importance of and technics for measuring the functional and relatively intangible instructional outcomes as well as the more tangible and easily measurable knowledges and skills. Wood and Haefner (56) wrote on the measurement and guidance of individual growth in a semi-popular style presumably based on the Lorge and Flesch readability concepts. Ross (41) produced a revision of his textbook in educational measurement shortly before his death. Blair (6) gave detailed attention to the diagnostic and remedial uses of standardized and nonstandardized instructional and practice tests in the classroom.

Wrinkle (60) made a significant addition to the voluminous literature on teacher marking of pupil progress in his report of the extended study of this question at the Colorado State College of Education. Methods tried experimentally were: (a) manipulating the symbols, (b) supplementing the symbols, (c) parent-teacher conferences, (d) informal letters to parents, (e) check forms, (f) pupil self-evaluation and reporting, and (g) parents' reports to the school.

Personality Measurement

Cattell (11) issued an extensive analytical and cross-sectional treatment of personality measurement which he expects to follow later with a developmental treatment of the hereditary, environmental, and somatic factors influencing personality. Bell (5) wrote a comprehensive treatment of projective technics which included not only the most widely used methods but also such varied and relatively little used procedures as the *Tautophone Test*, cloud pictures, expressive movement, finger painting, voice and speech, and the psychodrama. Frank (22) and Machover (33) also wrote on projective methods. Specialized treatments of the *Thematic Apperception Test* were issued by Stein (46) and Tomkins (52). Rapaport and others (38) wrote an extensive rationale of projective technics, with particularly extensive treatments of the *Rorschach* and the *Thematic Apperception Test*. The same authors (39) also issued a manual on the clinical use of projective technics as diagnostic aids.

Armed Forces Testing

Altho Flanagan and his collaborators dealt extensively in the December 1948 special issue of the REVIEW with testing in the armed forces, the most significant readily available sources appear to merit mention here. Four reports of Army Air Force testing procedures were edited by Deemer (19) on records, analysis, and test procedures, by Davis (18) on the qualifying examination, by Guilford and Lacey (26) on printed classification tests, and by Gibson (24) on motion picture testing. Davis (17) also discussed the selection and classification technics calculated to utilize human talent effectively in the armed services, and Stuit (49) edited the report on test development and personnel research technics in the Bureau of Naval Personnel.

The Assessment Staff of the U. S. Office of Strategic Services issued an extensive report (53) of procedures used in assessing 5391 persons between December 1943 and August 1945 for varied types of espionage and related responsibilities. As the diversity of responsibilities the selectees would assume was great, and as changes in war theaters often made necessary the assignment of selectees to duties for which they had not been recruited, assessed, or trained, usual methods of job selection were inappropriate. The OSS developed global assessment procedures which represent the "first attempt in America to design and carry out selection procedures in conformity with so-called organismic (Gestalt) principles." (53, p. 3) The tests and technics used in the assessment situations included some standardized instruments, some instruments constructed to meet definite needs, and many new technics in the fields of intelligence and aptitude, achievement of proficiency, and personality measurement.

Projected Publications

Several significant publications scheduled for issuance during 1950 have been announced. The two which involve wide participation by specialists in measurement and evaluation are discussed here, and the others will be mentioned in later sections of this chapter.

A.C.E. Measurement Book Project

A cooperatively planned and prosecuted contribution to achievement measurement is the book to result from what is known as the measurement book project of the American Council on Education. This is considered to be an expansion and up-to-date treatment of educational measurement comparable to the Hawkes, Lindquist, and Mann (27) cooperative book project which culminated in the 1936 publication. The tentative outline of the new book shows three major divisions: (a) the functions of measurement in education, (b) test construction, and (c) test theory. Part I

is expected to deal with the functions of measurement in the facilitation of learning, in improving the content, organization, supervision, and administration of instruction, in counseling, and in educational placement. Scheduled for Part II are chapters on preliminary considerations, planning the test, item writing, experimental tryout of the test, analysis of the tryout data and revision of the test, administration and scoring, reproducing the test, the performance test, and the essay test. Part III will probably include such chapters on test theory as the fundamental nature of measurement; reliability; validity; units, scales, and norms; and batteries and profiles.

Cooperative Study of Secondary-School Standards

The widely used 1940 materials for evaluation of secondary schools (13, 14) have been revised and tried out extensively in representative schools since approximately January 1, 1948.¹ Publication of the new materials is planned for March or April 1950. The new criteria differ significantly from the 1940 criteria in three major respects: (a) The evaluation of the objectives of the school will be based more on the needs of pupils and correspondingly less on its philosophy; (b) the evaluation of the educational program will include extracurriculum activities, the program of studies, and sixteen subject areas as replacements for the curriculum, extracurriculum activities, outcomes, and instruction; and (c) the summary of individual evaluations and the graphic report will be simplified and the "educational temperatures" method of summarizing results will be abandoned.²

Test Publishing and Research

The rate of production of new commercially available tests was understandably retarded during the war years, altho many tests for restricted use were prepared by various branches of the armed services. The period since 1945 has been marked by the issuance of many new tests and of revisions of previously published tests, by the merging of several test publishing agencies, and by an increasing attention to the proper use of test results.

Educational Testing Service

The Cooperative Test Service of the American Council on Education, the College Entrance Examination Board, and the Graduate Record Office

¹ The source of this statement is the minutes of the General Committee Meeting of the Cooperative Study of Secondary School Standards, Ann Arbor, September 1-3, 1949. (Jessen, Carl A., secretary.)

² Verification for this statement was obtained by the authors from David Segel and Carl A. Jessen in a letter dated November 30, 1949.

of the Carnegie Foundation for the Advancement of Teaching merged on January 1, 1948, to form the Educational Testing Service (12). The *National Teacher Examinations* and the *A.C.E. Psychological Examinations* were also included. Recently announced additions to the offerings are the college ability and proficiency tests (21), designed for predicting success in further college work of students who have completed some undergraduate study, and the *Pre-Engineering Inventory* (31), to be handled by the College Board division of the Educational Testing Service.

Use and Availability of Tests

The vast scale on which various tests are administered to persons of all ages and in all walks of life is shown by the summary presented by Reavis (40) of Educational Records Bureau estimates for 1944 testing. It was estimated that approximately 60 million tests were administered to about 20 million persons in the United States during that year. Although a large proportion of the tests were used for armed services and civil service testing, it is believed that about 26 million tests were administered to some 11 million persons in colleges, business firms, and the offices of personnel consultants.

Woodruff and Pritchard (57) indicated the wide variety of tests currently available from seventy-four publishers in 1948. They reported that their files included 1080 tests, of which 228 were in the areas of intelligence, aptitudes, and readiness, 716 were achievement tests of various types, and 90 were instruments designed to measure various aspects of personality and adjustment. Of the major subject fields into which the achievement tests were classified, English language and grammar, mathematics, reading, the social studies, science, and the foreign languages ranked in that order from high to low in the number available. Wrightstone (59) discussed the recent development and present status of aptitudes and achievement measurement and elsewhere (58) gave a running summary of the best new tests of aptitudes, readiness, achievement, work-study skills, critical thinking, and emotional and social adjustment.

Michaelis (34), surveying evaluation practices in sixty-eight city-school systems, found that fifty-two employed a director of evaluation, that fifty-four had cumulative record systems, that one-fifth of the total had a teacher's guide or handbook on evaluation, that all sixty-eight used tests, and that at least half made use each of interviews, case studies, case conferences, observation, group discussion, and anecdotal records in pupil evaluation.

Attacking a problem of great importance, Yale for the Science Research Associates, Durost for the World Book Company, and Bennett and Seashore for the Psychological Corporation (61) discussed precautions taken in test distribution as safeguards against both intentional and unintentional misuse of tests and misinterpretation of results.

Adkins (2) presented two lists of needed research on examining

devices, one based primarily on subprofessional and professional positions for which paper-and-pencil tests of basic abilities are desired, and the other consisting of various intangible abilities and preferences for which written tests would probably not be appropriate. She pointed out that the U. S. Civil Service Commission is interested in cooperating with universities in the prosecution of such studies.

Trends in Educational and Psychological Measurement

Certain trends in educational and psychological measurement appear to be in process as the second half-century of development in this area gets under way. These range from some which are only now emerging to others which have become well established over a period of years. To be considered here are a few of the general trends which the authors believe to merit mention. They supplement the trends in publication of books and in publication of tests and use of test results which are noted in preceding sections of this chapter.

One trend is found in the increasingly comprehensive appraisal of the individual child. Published evaluation programs of various schools have recently made this trend clearly evident. The schools are increasingly making use of cumulative records. Furthermore, statewide testing programs bear out this conclusion in their recent practices. Olson (36) recently embodied the basic principles of organismic or Gestalt psychology, which is represented by this trend in measurement practices, in his treatment of child development. The Office of Strategic Service employed a similar approach at a high level in its *Assessment of Men* (53).

A second trend which may well be a corollary of the first has appeared in the increasing use of multiple-aptitude tests, particularly at the secondary-school level, to supplement or even to replace general intelligence tests. The impetus for the use of these multiple-aptitude tests has come in part from the increasing emphasis in child study upon knowing the whole child and in part from the search for an aptitude test battery by the factor analysts of the War and Navy Departments during World War II and by civilian psychologists. The basic psychological principles upon which these tests rest have been brought together by Segel (44).

A third trend which may also be a corollary of the first is that embodied in the tests of general educational development, which are based on logical analysis rather than upon factor analysis procedures. First appearing in widely available form were the *U. S. Armed Forces Institute Tests of General Educational Development* (20) and the *Iowa Tests of Educational Development* at the college and high-school levels. The *Iowa Every-Pupil Tests of Basic Skills*, recently published in a single-booklet integrated edition, appear to embody this trend for the elementary-school grades. These tests place major emphasis upon the functional rather than upon the formal subject-centered aspects of learning.

A fourth trend is found in the increasing number of states inaugurating statewide testing programs which furnish consultative and field services in addition to basic testing services. This trend is doubtless the result of an increasing awareness of need for guidance in the measurement field by local-school officials, the increasing acceptance of responsibility for the provision of such guidance by state universities and state departments of education, and the fact that the testing program and use of the results should conform to the needs of each individual school system. Some of the states exemplifying this trend are Connecticut, Florida, Georgia, Illinois, Indiana, Iowa, Kentucky, Maryland, Michigan, Minnesota, Montana, New Hampshire, and South Carolina. Segel (45) issued a circular describing briefly these and other statewide programs.

A fifth trend which might almost be designated as a new approach in measurement is found in the greatly expanded research on the measurement of group and individual social status. The greatest impetus to the measurement of the social status of groups and individuals has come from the investigations made under University of Chicago sponsorship in various communities. Hollingshead's report (29) and the culminating book by Warner, Meeker, and Eells (54) described this approach to the measurement of social status.

A sixth trend, closely related to and probably a corollary of the fifth, is found in the increasing attention to research in the measurement of group dynamics, with particular attention to the participation of the individual and his contribution to group welfare. Olson (36) described the measurement of social participation by individuals. Further attention to group dynamics is found in the book by Bales (4), which deals with the analysis of interaction processes in small groups.

A seventh trend is found in the increasing attention to unstructured personality measurement and the development of technics for evaluating behavior in a wide variety of unstructured situations (5). Paralleling this emphasis upon projective technics is the tendency away from the uncritical acceptance of results from adjustment inventories and interest inventories. Demands for satisfactory demonstrations of validity are being applied increasingly both to structured inventories and to the projective methods.

The resultant of these various recent trends in educational and psychological measurement seems to the authors to have major implications. The often-mentioned collaboration of a test technician and subjectmatter specialists seems no longer to be completely adequate in itself. With the present emphasis upon areas of experience and the dynamics of behavior rather than solely upon subjectmatter, such a merger of talents appears to be somewhat lacking in vitality.

Certain premises seem to be widely accepted today. One is that the modern teacher must be concerned with the basically important and functional learning outcomes, usually far less tangible and less easily measurable than are knowledges and skills, which were largely disregarded

some years ago both by teacher and tester. Another is that the whole child must be evaluated in the dynamic social situations in which he inevitably finds himself. The demand, therefore, seems to be for an approach to evaluation which is founded upon a sound understanding of child growth and development, the nature of learning, and individual, group, and trait differences. The measurement specialist may well be faced with the necessity of meeting this type of demand. It seems likely to entail a broader equipment of testing technics and a greater use of the principles of expanding social and educational psychology than were demanded during the first fifty years of development in educational and psychological testing.

Measurement Activities of AERA Members

Scates (42), reporting on his 1948 survey of AERA membership, pointed out that "we are a technical group, concerned largely with measuring and counting, and engaging in that research which affords us an opportunity to do these things" (42, p. 137). One hundred forty-four members, more than a third of those responding to a question concerning research interests, listed appraisal, measurement, and test construction. This was the field named by the largest number of members. Of the 525 members who responded to a question concerning their special research skills, 415 reported competence in the statistical theory of measurement equivalent to a third semester of statistical training and 360 reported reasonably technical skill in objective test construction. Of the 607 members who reported their research activities, 413 indicated that they spent at least 10 percent of the working time on research involving the collection and analysis of data and the construction of new research instruments and 188 specified the teaching of research courses in statistical methods, statistical methods of test construction, and several variations apparently less closely allied to measurement as part of their regular duties.

Further evidence concerning the measurement activities of AERA members was obtained by one of the authors, who cross-checked the AERA membership as listed in the December 1948 REVIEW with Buros' index of reviewers, test and book authors, and other persons who variously contributed to the tests and books dealt with in his *Third Mental Measurements Yearbook* (8). Buros drew 77 of his 320 reviewers, or approximately 24 percent, from Association membership. Of the 596 AERA members, 214, or 36 percent, were listed in his index for one or more reasons. The 77 who each wrote one or more reviews constituted 13 percent of the AERA members.

Bibliography

1. ADKINS, DOROTHY C. *Construction and Analysis of Achievement Tests*. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 292 p.
2. ADKINS, DOROTHY C. "Needed Research on Examining Devices." *American Psychologist* 3: 104-106; March 1948.
3. ANDERSON, HOWARD R., and LINDQUIST, EVERET F. *Selected Test Items in American History*. Second edition. National Council for the Social Studies, Bulletin No. 6, Revised. Washington, D. C.: National Education Association, 1949. 113 p.
4. BALES, ROBERT F. *Interaction Process Analysis*. Cambridge, Massachusetts: Addison-Wesley Press, 1949. 203 p.
5. BELL, JOHN E. *Projective Techniques*. New York: Longmans, Green and Company, 1948. 533 p.
6. BLAIR, GLENN M. *Diagnostic and Remedial Teaching in the Secondary School*. New York: The Macmillan Company, 1946. 422 p.
7. BROWNELL, WILLIAM A., chairman. *The Measurement of Understanding*. Forty-Fifth Yearbook of the National Society for the Study of Education, Part I. Chicago: University of Chicago Press, 1946. 338 p.
8. BUROS, OSCAR K., editor. *The Third Mental Measurements Yearbook*. New Brunswick, New Jersey: Rutgers University Press, 1949. 1047 p.
9. BURT, CYRIL L. *Handbook of Tests for Use in Schools*. Revised edition. New York: Staples Press, 1948. 112 p.
10. BURT, CYRIL L. *Mental and Scholastic Tests*. Second edition. New York: Staples Press, 1947. 432 p.
11. CATTELL, RAYMOND B. *Description and Measurement of Personality*. Yonkers, New York: World Book Company, 1946. 602 p.
12. CHAUNCEY, HENRY. "Testing Offices of Three Organizations Merge." *College Board Review* 1: 36-37; Winter 1947.
13. COOPERATIVE STUDY OF SECONDARY SCHOOL STANDARDS. *Evaluation of Secondary Schools: General Report*. Washington, D. C.: the Study, 1939. 526 p.
14. COOPERATIVE STUDY OF SECONDARY SCHOOL STANDARDS. *Evaluative Criteria and Educational Temperatures*. 1940 edition. Washington, D. C.: the Study, 1939. 175 p.
15. CRAWFORD, ALBERT B., and BURNHAM, PAUL S. *Forecasting College Achievement*. New Haven, Connecticut: Yale University Press, 1946. 291 p.
16. CRONBACH, LEE J. *Essentials of Psychological Testing*. New York: Harper and Brothers, 1949. 475 p.
17. DAVIS, FREDERICK B. *Utilizing Human Talent: Armed Services Selection and Classification Procedures*. Commission on Implication of Armed Services Educational Programs. Washington, D. C.: American Council on Education, 1947. 86 p.
18. DAVIS, FREDERICK B., editor. *The AAF Qualifying Examination*. AAF Aviation Psychology Program Research Reports, No. 6. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 266 p.
19. DEEMER, WALTER L., JR., editor. *Records, Analysis, and Test Procedures*. AAF Aviation Psychology Program Research Reports, No. 18. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 621 p.
20. DETCHEN, LILY. "The United States Armed Forces Institute Examinations." *Educational Record* 28: 463-86; October 1947.
21. EDUCATIONAL TESTING SERVICE. "The New Intermediate Tests for College Students." *School and Society* 70: 332-33; November 19, 1949.
22. FRANK, LAWRENCE K. *Projective Methods*. Springfield, Illinois: Charles C. Thomas, 1948. 86 p.
23. FROELICH, CLIFFORD P., and BENSON, ARTHUR L. *Guidance Testing*. Chicago: Science Research Associates, 1948. 104 p.
24. GIBSON, JAMES J., editor. *Motion Picture Testing and Research*. AAF Aviation Psychology Program Research Reports, No. 7. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 267 p.
25. GOODENOUGH, FLORENCE L. *Mental Testing: Its History, Principles, and Applications*. New York: Rinehart and Company, 1949. 609 p.

26. GUILFORD, JOY P., and LACEY, JOHN I., editors. *Printed Classification Tests*. AAF Aviation Psychology Program Research Reports, No. 5. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 919 p.
27. HAWKES, HERBERT E.; LINDQUIST, EVERET F.; and MANN, CHARLES R., editors. *The Construction and Use of Achievement Examinations*. Boston: Houghton Mifflin Company, 1936. 497 p.
28. HILDRETH, GERTRUDE. *A Bibliography of Mental Tests and Rating Scales—1945 Supplement*. New York: Psychological Corporation, 1946. 86 p.
29. HOLLINGSHEAD, AUGUST DE B. *Elmtown's Youth: The Impact of Social Classes on Adolescents*. New York: John Wiley and Sons, 1949. 480 p.
30. INTERNATIONAL BUSINESS MACHINES CORPORATION, DEPARTMENT OF EDUCATION. *Methods of Adapting Tests for Scoring by the IBM Electric Test Scoring Machine*. New York: the Corporation, 1947. 22 p.
31. JOHNSON, A. PEMBERTON. "CEEB, PEI Merge Programs." *College Board Review* No. 9: 105-106; December 1949.
32. KRAKOWER, HYMAN. *Tests and Measurements Applied to Nursing Education*. New York: G. P. Putnam's Sons, 1949. 179 p.
33. MACHOVER, KAREN. *Personality Projection in the Drawing of the Human Figure*. Springfield, Illinois: Charles C. Thomas, 1949. 194 p.
34. MICHAELIS, JOHN U. "Current Practices in Evaluation in City School Systems." *Educational and Psychological Measurement* 9: 15-22; Spring 1949.
35. MURSELL, JAMES L. *Psychological Testing*. Second edition. New York: Longmans, Green and Company, 1949. 488 p.
36. OLSON, WILLARD C. *Child Development*. Boston: D. C. Heath and Company, 1949. 417 p.
37. PEASE, KATHARINE. *Machine Computation of Elementary Statistics*. New York: Chartwell House, 1949. 208 p.
38. RAPAPORT, DAVID, and OTHERS. *Diagnostic Psychological Testing: The Theory, Statistical Evaluation, and Diagnostic Application of a Battery of Tests*. Volume II. Chicago: Year Book Publishers, 1946. 516 p.
39. RAPAPORT, DAVID, and OTHERS. *Manual of Diagnostic Psychological Testing: II, Diagnostic Testing of Personality and Ideational Content*. New York: Josiah Macy, Jr., Foundation, 1946. 105 p.
40. REAVIS, WILLIAM C. "Testing Is Big Business." *School Review* 55: 259-60; May 1947.
41. ROSS, CLAY C. *Measurement in Today's Schools*. Second edition. New York: Prentice-Hall, 1947. 551 p.
42. SCATES, DOUGLAS E. "Analysis of a Professional Organization: The American Educational Research Association in 1948." *Growing Points in Educational Research*. 1949 Official Report of the American Educational Research Association. Washington, D. C.: the Association, a department of the National Education Association, 1949. p. 111-42.
43. SCATES, DOUGLAS E. "Fifty Years of Objective Measurement and Research in Education." *Journal of Educational Research* 41: 241-64; December 1947.
44. SEGEL, DAVID. *Intellectual Abilities in the Adolescent Period*. U. S. Office of Education Bulletin, 1948, No. 6. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1948. 41 p.
45. SEGEL, DAVID. *Status of State-Wide Testing Programs*. U. S. Office of Education Circular, No. 320. Washington, D. C.: U. S. Office of Education, 1950.
46. STEIN, MORRIS I. *The Thematic Apperception Test*. Cambridge, Massachusetts: Addison-Wesley Press, 1948. 95 p.
47. STEPHENSON, WILLIAM. *Testing School Children: An Essay in Educational and Social Psychology*. New York: Longmans, Green and Company, 1949. 127 p.
48. STUIT, DEWEY B., chairman. *Predicting Success in Professional Schools*. American Council on Education Studies, Series VI. Washington, D. C.: the Council, 1949. 187 p.
49. STUIT, DEWEY B., editor. *Personnel Research and Test Development in the Bureau of Naval Personnel*. Princeton, New Jersey: Princeton University Press, 1947. 513 p.
50. SUPER, DONALD E. *Appraising Vocational Fitness by Means of Psychological Tests*. New York: Harper and Brothers, 1949. 727 p.
51. THORNDIKE, ROBERT L. *Personnel Selection: Test and Measurement Techniques*. New York: John Wiley and Sons, 1949. 358 p.

52. TOMKINS, SILVAN S. *The Thematic Apperception Test: the Theory and Technic of Interpretation*. New York: Grune and Stratton, 1947. 297 p.
53. U. S. OFFICE OF STRATEGIC SERVICES, ASSESSMENT STAFF. *Assessment of Men*. New York: Rinehart and Company, 1948. 541 p.
54. WARNER, W. LLOYD; MEEKER, MARCHIA; and EELLS, KENNETH W. *Social Class in America: A Manual of Procedure for the Measurement of Social Status*. Chicago: Science Research Associates, 1949. 274 p.
55. WEITZMAN, ELLIS, and MCNAMARA, WALTER J. *Constructing Classroom Examinations: A Guide for Teachers*. Chicago: Science Research Associates, 1949. 153 p.
56. WOOD, BEN D., and HAEFNER, RALPH. *Measuring and Guiding Individual Growth*. New York: Silver Burdett Company, 1948. 535 p.
57. WOODRUFF, ASAHEL D., and PRITCHARD, MARALYN W. "Some Trends in the Development of Psychological Tests." *Educational and Psychological Measurement* 9: 105-108; Spring 1949.
58. WRIGHTSTONE, J. WAYNE. "Broader Scope for Tests and Measurements." *Nation's Schools* 43: 39-41; March 1949.
59. WRIGHTSTONE, J. WAYNE. "Frontiers in Educational Research in the Measurement of Aptitudes and Achievement." *Journal of Educational Research* 40: 389-96; January 1947.
60. WRINKLE, WILLIAM L. *Improving Marking and Reporting Practices*. New York: Rinehart and Company, 1947. 120 p.
61. YALE, JOHN R., and OTHERS. "Checking the Qualifications of Users of Tests." *American Psychologist* 1: 353-57; August 1946.

CHAPTER II

Construction and Educational Significance of Intelligence Tests

ETHEL L. CORNELL and ANNETTE GILLETTE¹

General Trends

New Developments in Intelligence Testing

DEVELOPMENTS of measurement technics in the last three years have in one sense been rapid, but at the same time they have left so-called tests of general intelligence not much advanced beyond their status three years ago. That is to say, the direction which research in this field has taken has been away from efforts at over-all tests and toward a breakdown into more adequate tests of "factors" or "functions" or "aptitudes." Davis (12) concluded from a study of results of testing in the armed services that combinations of highly specialized aptitude tests are more effective for guidance purposes than are tests of general intelligence or of general learning ability.

New tests published during the period include the civilian edition of the *Army General Classification Test* (45, 58), the *Differential Aptitude Tests* (3), two shorter forms of the *Primary Mental Abilities Tests* (55), a battery adapted from War Department tests for high-school use by Segel (46), and a wide-range picture vocabulary test (1). While these tests have much the same purpose as general intelligence tests of the past, they are designed to measure relatively independent traits. The traits included in various batteries are not the same, but all include both verbal and nonverbal factors which test a wider range of abilities than most intelligence tests of the past. The problem of finding adequate criteria and of determining the predictive importance of factors not closely related to academic success is still unsolved.

Factors, Traits, Types of Content

Factor analysis continues to show a few factors that recur in factoring various tests and in using various methods of analysis, but there is still a great deal of confusion. The number of different factors found in any given test, the degree to which factor loadings are found constant for the same type of test by different investigators, the number of non-intellective factors that may influence any test at different times, and the

¹ The section on *Applications of Intelligence Tests* was prepared by Dr. Gillette.

relative complexity of factors at different age levels are a few of the unsolved problems to which attention is being called.

Tests that might be subjectively regarded as similar are not always found to be related factorially (13). Different methods of factor analysis may yield different factors. Wenger, Holzinger, and Harman (62) compared bi-factor and multiple-factor methods and found that three uncorrelated group factors revealed by the bi-factor method were resolved into at least four by a multiple-factor method. Fruchter (16) identified at least two factors in verbal fluency: *F*, related to the flow of responses, and *S*, related to the selection of responses required for the solution of the problem.

Shaw (47) comparing *Primary Mental Abilities* tests with tests of achievement at the high-school level, found that the tests of verbal meaning and reasoning were the only ones showing much correlation with achievement tests and that the optimum combination of all the *Primary Mental Abilities* tests accounted for about 20 percent of the variance in reading rate but for 61 percent of the variance in vocabulary.

Tests of intelligence, to give the obverse of the picture, may be more heavily loaded with nonintellective factors than with intellective ones. Jastak (28) indicated that "intelligence" accounts for only 20 to 25 percent of the variance of any one test, and that the remaining variance must be accounted for by factors independent of intellectual level.

There is quite general agreement now among factor analysts that most tests contain a general factor. Jastak (29) stated his belief that *g* (if properly measured) may afford the best reference point for the evaluation of group factors. Group factors might be found to represent personality traits which are independent of intellectual capacity.

A critical study of the assumptions underlying test construction was made by Loevinger (32). She developed the concept that there are three major problems of test construction still unsolved, whose solution is presupposed or assumed in factor analysis. The three problems relate to self-consistency of tests (reliability), choice of items, and the unit of measurement. She also showed that the formulas accepted and used in establishing reliability, selecting items, and scaling tests are dependent on "assumptions which are remote from if not contradictory to the realities of the clinical situation in testing." Some of the disappointing results of factor analysis are probably due to the fact that the basic test data do not meet the technical requirements necessary for the underlying assumptions of factor analysis.

The question of an adequate criterion by which to judge intelligence tests is another problem that has received recent attention. Tyler (57) pointed out the changing emphasis in educational evaluation with respect to the method of formulating the problem. He indicated that finding measures of individual characteristics which will predict success in schools *as they are* has given way to finding measurable abilities which can be developed into socially and personally valuable behavior if school programs are planned to capitalize on these abilities.

A number of studies sponsored by Havighurst and Davis at the University of Chicago have made an approach to this problem by trying to isolate the effect of cultural status on intelligence tests (9, 10, 11, 14, 21, 22, 38, 44). That the differences in test results consistently shown between higher and lower cultural class groups is due at least in part to the "socio-economic bias" of present tests is strongly suggested by the work of Davis (9), who changed certain test items to eliminate the cultural loading of the content in such manner that the essential problems appeared to be unchanged. This seems to be a promising approach. If Davis' findings are corroborated—that tests including syllogisms, logical classification, inductive reasoning, arithmetic reasoning, and problems of imaginative insight, when couched in terms commonly understood by different social-class groups, show no significant differences between the groups—it seems that educational method is in need of a major reorientation.

Considering the unsolved problems in test construction and factor analysis and the importance of pattern and organization revealed by factor analysis, the contribution of experimental studies in concept formation to the problem of defining "intelligence" should not be ignored. Heidebreder (23), continuing her studies on the attainment of concepts by using card-sorting tests, found that there is a hierarchical order of concept formation which is determined by factors within the organism, but that this order can be varied by the amount of "situational support" (perceptual or semantic cues contained in the test cards) toward one or another mode of conceptual behavior. The order of concept formation was found to be the same for college students and adults with no more than elementary-school education. Differences between these groups were shown in the speed of attaining a given concept, in the facility of shifting from one mode to another (in classifying cards according to different possibilities of classification), in the ability to entertain several possibilities of classification at once, and in maintaining a new set once it was adopted. Similar test approaches were made by Welch (61) and by Weinberg in France (60). Tests of this type seem to the present reviewer to have important implications for teaching and learning.

Growth and Development

Factorial methods have also been used to study growth trends. Whether factors become more or less complex with increasing age and whether the same factors are found at various ages are problems as yet without conclusive answers. It has generally been thought that abilities are less differentiated at younger ages and that they break down in adolescence into relatively independent factors (18). Recent evidence has raised some doubt about this hypothesis. Clark (7) found that about the same number of factors were required to account for the correlation matrix at each

age level from Grades I to XII. Chen and Chow (6), however, concluded that the factor pattern tended to become simpler with increasing age and that the factor loading of g increased from 0.42 in the primary grades to 0.64 at the senior-high-school level. Swineford's results (50) seem to corroborate the findings of Chen and Chow. Her results also supported the hypothesis that factors remain the same from age to age, at least in the narrow range from Grade VII to Grade IX.

Thorndike (53) measured gains of 1000 students on *American Council on Education Psychological Examinations* over a nine-year period, from age 13½ to age 20. The growth curve became flat at 21½ years or at 25 9/12 years, according to the method of extrapolating. To what extent this would be true for a group not pursuing formal education thruout these years is an open question. Vernon (59) found a deterioration in educational attainment but continued growth in mechanical and spatial tests for Royal Army and Navy recruits after age fourteen with no further education.

The relative status of Terman's gifted children twenty years after their original selection was reported (52), using as a criterion the vocabulary test standards developed by Thorndike and Gallup on a Gallup poll. Terman's group, originally the top 1 percent of their age group, now range over the top half of the "voting public." About half are in the top 5 percent and most are in the top 25 percent.

Applications of Intelligence Tests

Intelligence Tests and Educational Achievement

Schafer and Leitch (41) reported that psychologists were able to indicate the degree of maladjustment for twenty-two nursery-school children thru results of psychological tests. Hobson (26) studied the school achievement over a period of ten years of a group of children who were selected on the basis of mental age and admitted to kindergarten and first grade from three to nine months younger than the chronological age required for entrance in Brookline, Massachusetts. At all grade levels except kindergarten the underage group had the better marks and they attained achievement test averages which were from two to seven months higher than those of the older group in the same grade.

Thomas (51) compared reading comprehension scores on the *Stanford Achievement Test* for sixth-grade pupils ($N = 2918$) in three successive years with their mental ages obtained on the *Otis Quick-Scoring Test of Mental Ability, Beta*. About 13 percent were found to have reading ages a year or more below the level expected for their mental ages.

Kvaraceus and Lanigan (31), using results for twenty-seven junior-high-school pupils tested at half-year intervals over a period of two years, found that scores on the *Otis Quick-Scoring Mental Ability Test, Beta*

correlated .78 to .59 with scores on various parts of the *Iowa Every-Pupil Tests of Basic Skills*, the lowest correlation being with arithmetic and the highest with reading comprehension.

Bolton (4) studied the relation between marks for two semesters and scores on several group intelligence tests for a group of 212 seniors in one high school. Using two tests improved the ability to predict marks, but the standard error of estimate was still too high to consider the intelligence test ratings as more than suggestive of academic success. It was concluded that the *California Non-Language Intelligence Test* combined with one of the verbal intelligence tests facilitated educational guidance.

For 100 boys of high-school age in a correctional institution, Cofer and Biegel (8) found that the *Kent Rapid Screening Test* was not as closely associated with educational achievement as with scores on the *Buck Time Appreciation Test*.

For a group of 1681 college freshmen Wheeler and Wheeler (63) reported a high relation between reading scores and the *L* score on the *American Council on Education Psychological Examination* ($r = .70$), and a low relation between reading scores and the *Q* score ($r = .36$). The authors concluded that the A.C.E. test is materially influenced by reading efficiency, a factor which should be considered in interpretation and use of results.

For two classes of a teachers college Haas (19) discovered that rank in high school was more predictive of academic success than percentile rank on the *Henmon-Nelson Test of Mental Ability*. The correlations of the latter with honor points were .34 and .32 for the two classes. Rausch (40) related degrees of individual variation between quantitative and linguistic scores on the A.C.E. test to academic achievement based on marks for 1551 college students. The least variable group exceeded the middle and the most variable groups in scholastic achievement. The results would have been more conclusive if the author had given data to show whether the intelligence level of the three groups differed, as he did in presenting material on the *Cooperative English Test*. The author concluded that variability in the individual is not conducive to superior academic achievement, and suggested that averages of scores on separate tests have less predictive value than does knowledge of the individual's variability. Tilton (56), studying variability of I.Q.'s obtained from five general intelligence tests in relation to intelligence level, found a tendency for the bright pupils to be less variable and the dull to be more variable than the average, tho the difference was not significant.

Long and Perry (33) discussed the tests used as entrance examinations at the College of the City of New York, indicating changes which were made on the basis of ability to predict first-semester grades.

Using results from the *Scholastic Aptitude Test* of the College Entrance Examination Board for 5000 pupils, 600 of whom had been tested ten years or more before college entrance, Thorndike (54) concluded that

intelligence tests given one to three years before college entrance were as predictive of college board examination results as tests given at the same time. When more than one test had been given, a simple average gave the best prediction of later test scores and was more reliable than a single score.

Results reported in special fields are not entirely consistent. Havens (20) found that reading tests afforded a better prediction of law school grades than did either the A.C.E. test or high-school rank. McClanahan and Morgan (34) similarly found the A.C.E. test to be of little value in predicting freshmen grades for engineering students. However, Moser's results (36) for 550 high-school seniors who selected three occupations as possible future careers showed that the vocations requiring advanced professional training tended to be chosen by students of higher tested intelligence. In the field of art, Barrett (2), using high-school students, and Bottorf (5), with college students, found that intelligence was a factor in success in art courses.

Problems Relating to Test Administration

Super, Braasch, and Shay (49) concluded that commonly occurring disturbances in testing conditions do not affect test results of graduate students for the *Otis Quick-Scoring Mental Ability Test*, *Gamma*, and other tests.

A report (27) of retests of children from the ages of two to eighteen years of age, employing several individual and group mental tests, indicated decreasing correlations of I.Q.'s with longer intervals. Almost 60 percent changed fifteen or more I.Q. points, the greatest fluctuations occurring when there were unusual variations in disturbing or stabilizing factors in the child's environment. Hilden (24) followed thirty children as a part of a longitudinal study and found fluctuations of from seven to forty-six points in I.Q.'s.

Stalnaker and Stalnaker (48), in repeated testing with the *Scholastic Aptitude Test* of the College Entrance Examination Board, concluded that the higher score when the test is repeated is due to growth in the verbal factor rather than to familiarity with the test. Muntyan (37) found significant gains on identical and comparable forms of the A.C.E. test and the *Illinois High School Tests in Reading*. He suggested the use of separate norms for test and retest of high-school students.

MacPhee, Wright, and Cummings (35) indicated cautions as to the pitfalls of interpretation in using an abbreviated test.

Relation Between the School Program and Intelligence Level

Fulst (17) cited a case to illustrate the significant increases in reading skill, intelligence (as measured by the *A.C.E. Psychological Examination*),

and social acceptance which were brought about by furthering good human relationships for a group of seventh- and eighth-grade students.

Fattu (15) rated the rural schools from which ninth-grade students entered the University High School at Bloomington, Indiana, in physical aspects and in teaching and learning atmosphere and concluded that there was a definite relation between ratings of school adequacy and performance of pupils on objective tests. Orr (39) found that students of relatively low ability who were prepared in first-class high schools remained in college longer than did students of comparable ability who came from second- and third-class high schools.

Schmidt's study (42), of which a preliminary account was reported in the last REVIEW on this topic, was severely criticized by Kirk (30) on the grounds of inaccurate and misleading reporting of data, and also by Hill (25). Schmidt's reply (43) to Kirk does not answer his criticisms. Her implications for the education of retarded children should be of sufficient concern to educators to make an unbiased review of her original data a matter of importance.

Summary

In summary it may be said that theoretical questions concerning the purposes for which tests are constructed and the fundamental interpretations that may be legitimately made are occupying the attention of test makers and that test users might avoid some pitfalls if they would consider these questions more carefully in the planning of research and in the interpretation of their findings.

Bibliography

1. AMMONS, ROBERT B., and HUTH, RICHARD W. "The Full-Range Picture Vocabulary Test. I. Preliminary Scale." *Journal of Psychology* 28: 51-63; July 1949.
2. BARRETT, HARRY O. "Examination of Certain Standardized Art Tests To Determine Their Relation to Classroom Achievement and to Intelligence." *Journal of Educational Research* 42: 398-400; January 1949.
3. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *Differential Aptitude Tests*. New York: Psychological Corporation, 1947.
4. BOLTON, FLOYD B. "Value of Several Intelligence Tests for Predicting Scholastic Achievement." *Journal of Educational Research* 41: 133-38; October 1947.
5. BOTTORF, EDNA A. "Study Comparing Art Abilities and General Intelligence of College Students." *Journal of Educational Psychology* 37: 398-426; October 1946.
6. CHEN, TSO-YU L., and CHOW, HUAL-HSUL. "A Factor Study of a Test Battery at Different Educational Levels." *Pedagogical Seminary and Journal of Genetic Psychology* 73: 187-99; December 1948.
7. CLARK, WILLIS W. "The Differentiation of Mental Abilities at Various Age Levels." *Growing Points in Educational Research*. 1949 Official Report, American Educational Research Association. Washington, D.C.: the Association, a department of the National Education Association, 1949. p. 39-43.
8. COFER, CHARLES N., and BIEGEL, MARK M. "A Study of the Kent and Buck Screen Tests of Mental Ability in Relation to Otis and Stanford Achievement Test Scores." *Journal of Consulting Psychology* 12: 187-89; May-June 1948.

9. DAVIS, ALLISON. "Education and the Conservation of Human Resources." *Education and the General Welfare*. Official Report 1949 Convention, American Association of School Administrators. Washington, D. C.: the Association, a department of the National Education Association, 1949. p. 74-83.
10. DAVIS, ALLISON. *Social-Class Influence Upon Learning*. Cambridge, Massachusetts: Harvard University Press, 1948. 100 p.
11. DAVIS, ALLISON, and HAVIGHURST, ROBERT J. "Measurement of Mental Systems: Can Intelligence Be Measured?" *Scientific Monthly* 66: 301-16; April 1948.
12. DAVIS, FREDERICK B. *Utilizing Human Talent: Armed Services Selection and Classification Procedures*. Commission on Implications of Armed Services Educational Programs. Washington, D. C.: American Council on Education, 1947. 86 p.
13. DEJMEK, FRANK W. "A Study of Relationships Among Scores on the Minnesota Paper Form Board—Revised Series AA, the Non-Language Section of the California Test of Mental Maturity and Test 2 of the Non-Language Section of the California Test of Mental Maturity." *Journal of Educational Research* 42: 307-11; December 1948.
14. ELLS, KENNETH W. *Social Status Factors in Intelligence Test Items*. Doctor's thesis. Chicago: University of Chicago, 1948. 686 p.
15. FATTU, NICHOLAS A. "Some Relationships Between Indices of School Adequacy and Certain Measures of Pupil Achievement." *Indiana University School of Education Bulletin* 23: 21-32; November 1947.
16. FRUCHTER, BENJAMIN. "The Nature of Verbal Fluency." *Educational and Psychological Measurement* 8: 33-47; Spring 1948.
17. FULTS, ANNA C. "Improving Learning Through an Emphasis on Human Relations." *Childhood Education* 24: 305-307; March 1948.
18. GARRETT, HENRY E. "A Developmental Theory of Intelligence." *American Psychologist* 1: 372-78; September 1946.
19. HAAS, LEONARD. "Four Year Studies of the Freshman Classes of 1936 and 1940 at the Eau Claire State Teachers College." *Journal of Educational Research* 42: 54-61; September 1948.
20. HAVENS, VIRGINIA. "A Prediction of Law School Achievement from High-School Rank, Reading Test Score, Psychological Test Scores and Average Grade in Pre-Law Courses." *Journal of Educational Psychology* 39: 237-42; April 1948.
21. HAVIGHURST, ROBERT J. "Culture and the I.Q." *School Review* 57: 187-89; April 1949.
22. HAVIGHURST, ROBERT J., and BREEZE, FAY H. "Relation Between Ability and Social Status in a Midwestern Community. III. Primary Mental Abilities." *Journal of Educational Psychology* 38: 241-47; April 1947.
23. HEIDBREDER, EDNA F. "The Attainment of Concepts. VII. Conceptual Achievements During Card-Sorting. VIII. The Conceptualization of Verbally Indicated Instances." *Journal of Psychology* 27: 3-39, 263-309; January, April 1949.
24. HILDEN, ARNOLD H. "A Longitudinal Study of Intellectual Development." *Journal of Psychology* 28: 187-214; July 1949.
25. HILL, ARTHUR S. "Does Special Education Result in Improved Intelligence for the Slow Learner?" *Journal of Exceptional Children* 14: 207-13; April 1948.
26. HOBSON, JAMES R. "Mental Age as a Workable Criterion for School Admission." *Elementary School Journal* 48: 312-21; February 1948.
27. HONZIK, MARJORIE P.; MACFARLANE, JEAN W.; and ALLEN, LUCILLE. "The Stability of Mental Test Performance Between Two and Eighteen Years." *Journal of Experimental Education* 17: 309-24; December 1948.
28. JASTAK, JOSEPH. "A Plan for the Objective Measurement of Character." *Journal of Clinical Psychology* 4: 170-78; April 1948.
29. JASTAK, JOSEPH. "Problems of Psychometric Scatter Analysis." *Psychological Bulletin* 46: 177-97; May 1949.
30. KIRK, SAMUEL A. "An Evaluation of the Study by Bernadine G. Schmidt Entitled: 'Changes in Personal, Social and Intellectual Behavior of Children Originally Classified as Feeble-minded.'" *Psychological Bulletin* 45: 321-33; July 1948.
31. KVARACEUS, WILLIAM C., and LANIGAN, MARY A. "Pupil Performance on the Iowa Every-Pupil Test of Basic Skills Administered at Half-Year Intervals in the Junior High School." *Educational and Psychological Measurement* 8: 93-100; Spring 1948.
32. LOEVINGER, JANE. "A Systematic Approach to the Construction and Evaluation of Tests of Ability." *Psychological Monographs* 61, No. 4; 1947. 49 p.

33. LONG, LOUIS, and PERRY, JAMES D. "Entrance Examinations at the City College of New York." *Educational and Psychological Measurement* 7: 765-72; Winter 1947.
34. McCLANAHAN, WALTER R., and MORGAN, DAVID H. "Use of Standard Tests in Counseling Engineering Students in College." *Journal of Educational Psychology* 39: 491-501; December 1948.
35. MACPHEE, H. M.; WRIGHT, H. F.; and CUMMINGS, S. B., JR. "The Performance of Mentally Subnormal Rural Southern Negroes on the Verbal Scale of the Bellevue Intelligence Examination." *Journal of Social Psychology* 25: 217-29; May 1947.
36. MOSER, WILBUR E. "Vocational Preference as Related to Mental Ability." *Occupations* 27: 460-61; April 1949.
37. MUNTYAN, MILOSH. "A Study of the Re-Test Factor in the Illinois Statewide High School Testing Program." *Journal of Educational Research* 41: 183-92; November 1947.
38. MURRAY, WALTER J. *The Intelligence Test Performance and Negro Children of Different Social Classes*. Doctor's thesis. Chicago: University of Chicago, 1947. 128 p.
39. ORR, HARRIET KNIGHT. "A Comparison of the Records Made in College by Students from Fully Accredited High Schools with Those of Students Having Equivalent Ability, from Second and Third Class High Schools." *Journal of Educational Research* 42: 353-64; January 1949.
40. RAUSCH, OSCAR P. "The Effects of Individual Variability on Achievement." *Journal of Educational Psychology* 39: 469-78; December 1948.
41. SCHAFER, SARAH, and LEITCH, MARY. "An Exploratory Study of the Usefulness of a Battery of Psychological Tests with Nursery School Children." *American Journal of Psychiatry* 104: 647-52; April 1948.
42. SCHMIDT, BERNADINE G. "Changes in Personal, Social and Intellectual Behavior of Children Originally Classified as Feeble-minded." *Psychological Monographs* 60, No. 5; 1946. 144 p.
43. SCHMIDT, BERNADINE G. "A Reply." *Psychological Bulletin* 45: 334-43; July 1948.
44. SCHULMAN, MARY JEAN, and HAVIGHURST, ROBERT J. "Relations Between Ability and Social Status in a Midwestern Community. IV. Size of Vocabulary." *Journal of Educational Psychology* 38: 437-42; November 1947.
45. SCIENCE RESEARCH ASSOCIATES. *Army General Classification Test, First Civilian Edition*. Chicago: Science Research Associates, 1947. 18 p.
46. SEGEL, DAVID. "The Validity of a Multiple Aptitude Test at the Secondary School Level." *Educational and Psychological Measurement* 7: 695-705; Winter 1947.
47. SHAW, DUANE C. "A Study of the Relationships Between Thurstone Primary Mental Abilities and High School Achievement." *Journal of Educational Psychology* 40: 239-49; April 1949.
48. STALNAKER, RUTH C., and STALNAKER, JOHN M. "The Effect on a Candidate's Score of Repeating the Scholastic Aptitude Test at the College Entrance Examination Board." *Educational and Psychological Measurement* 6: 495-503; Winter 1946.
49. SUPER, DONALD E.; BRAASCH, WILLIAM F., JR.; and SHAY, JOSEPH B. "The Effect of Distractions on Test Results." *Journal of Educational Psychology* 38: 373-77; October 1947.
50. SWINEFORD, FRANCES. "Growth in the General and Verbal Bi-Factors from Grade VII to Grade IX." *Journal of Educational Psychology* 38: 257-72; May 1947.
51. THOMAS, GEORGE I. "A Study of Reading Achievement in Terms of Mental Ability." *Elementary School Journal* 47: 28-33; September 1946.
52. THORNDIKE, ROBERT L. "An Evaluation of the Adult Intellectual Status of Terman's Gifted Children." *Pedagogical Seminary and Journal of Genetic Psychology* 72: 17-27; September 1948.
53. THORNDIKE, ROBERT L. "Growth of Intelligence During Adolescence." *Pedagogical Seminary and Journal of Genetic Psychology* 72: 11-15; September 1948.
54. THORNDIKE, ROBERT L. "The Prediction of Intelligence at College Entrance from Earlier Test." *Journal of Educational Psychology* 38: 129-48; March 1947.
55. THURSTONE, LOUIS L., and THURSTONE, THELMA G. *SRA Primary Mental Abilities, for Ages 7 to 11 and 11 to 17*. Chicago: Science Research Associates, 1947.

56. TILTON, J. WARREN. "The Relation Between IQ and Trait Difference as Measured by Group Intelligence Tests." *Journal of Educational Psychology* 38: 343-52; October 1947.
57. TYLER, RALPH W. "Educability and the Schools." *Elementary School Journal* 49: 200-12; December 1948.
58. U. S. ADJUTANT GENERAL'S OFFICE, PERSONNEL RESEARCH SECTION, STAFF. "The Army General Classification Test, with Special Reference to the Construction and Standardization of Forms 1a and 1b." *Journal of Educational Psychology* 38: 385-420; November 1947.
59. VERNON, PHILIP E. "Changes in Abilities from 14 to 20 Years." *Advancement of Science* 5: 138; July 1948.
60. WEINBERG, D. "Recherche sur une Épreuve d'Efficiencce Mentale en Fonction des Méthodes Successive ou Simultanée de Travail. Application au Diagnostic Individuel." *Travail Humain* 10: 68-83; 1947. (Abstract in *Psychological Abstracts* 22: 536; October 1948.)
61. WELCH, LIVINGSTON. "A Behaviorist Explanation of Concept Formation." *Pedagogical Seminary and Journal of Genetic Psychology* 47: 201-22; December 1947.
62. WENGER, MARION A.; HOLZINGER, KARL J.; and HARMAN, HARRY H. *The Estimation of Pupil Ability by Three Factorial Solutions*. University of California Publications in Psychology, Vol. 5, No. 8. Berkeley: University of California, 1948. 252 p.
63. WHEELER, LESTER R., and WHEELER, VIOLA D. "Relationship Between Reading Ability and Intelligence Among University Freshmen." *Journal of Educational Psychology* 40: 230-38; April 1949.

CHAPTER III

Construction and Educational Significance of Aptitude Tests

DEWEY B. STUIT

THE period covered by this REVIEW has seen a steady development of aptitude testing in education as well as in government, business, and industry. Carter reviewed research on essentially this topic in February 1947. The experience with aptitude tests during the war, the establishment of many Veterans Administration Guidance Centers, and the improved tests published during the period have all stimulated the wider use of aptitude tests. While the type of tests which appeared during this period is not radically new, the reports of research lead to the impression that better use is being made of old tests and that the newer tests reflect the experiences of educational personnel workers and of psychologists in business, industry, and the armed services. A new and comprehensive account of aptitude and other tests of value in vocational counseling was published by Super (47).

The term "aptitude test" is still being used in a variety of ways. In general it appears that the term is being applied primarily to tests which measure abilities or accomplishments which are not the *direct* result of specific environmental experiences and which are used to predict success at some future time (44). In effect this is the definition which was followed in selecting the materials to be included in this REVIEW. So-called intelligence and educational achievement tests and personality and interest inventories are excluded on the basis of this definition.

The most suitable method of classifying aptitude tests appears to be on the basis of the subjectmatter areas or occupational fields in which they are presumed to predict success. This type of classification is followed in this REVIEW.

Differential Aptitude Tests

The most recent development in the field of aptitude testing is that of the differential aptitude test battery. The fundamental idea in this type of testing stems from the work of the factor analysts, notably Thurstone's tests of primary mental abilities. The tests comprising a differential aptitude test battery presumably measure different facets of mind, thus revealing the individual's strengths and weaknesses in specific areas. While the tests recently published do not purport to measure the "pure" abilities represented in Thurstone's *Chicago Tests of Primary Mental*

Abilities, they nevertheless are designed to measure abilities which are not highly correlated. Such tests appear to provide the best answer to the often heard remark, "I want to take an aptitude test."

Test batteries resembling the differential aptitude test batteries recently published were used extensively during the war, notably by the Army Air Forces (19) and the Navy (46). These tests were peculiarly appropriate for use in the armed services because all able-bodied men had to be accepted for assignment and the problem became one of finding the best niche for the individual in question. It appears that in vocational guidance and in business and industrial personnel work similar needs exist.

It should be emphasized that a mere assembling of different tests does not constitute construction of a differential aptitude test battery. It must be shown that the tests comprising such a battery do in fact measure different abilities as evidenced by relatively low intercorrelations. Presumably different tests or combinations of tests can then be used to predict success in different subjectmatter fields or professional areas. To be useful in such a wide field of prediction means, of course, that a great deal of validation work must be done before such a battery can really be useful in vocational counseling. As will be seen in the paragraphs that follow, such validation studies are still rare.

During the last three years several differential aptitude tests have been published. Bennett, Seashore, and Wesman (4) devised the *Differential Aptitude Tests*, which consist of seven tests: verbal reasoning, numerical ability, abstract reasoning, space relations, mechanical reasoning, clerical speed and accuracy, and language usage (spelling and sentences). The tests are designed for use in Grades VIII to XII. There are two forms of each test except mechanical reasoning. The tests, other than clerical, are of the power rather than the speed type. Average reliability coefficients, with the exception of that for girls on mechanical reasoning, which is .71, range from .85 to .93. There are separate percentile norms for boys and girls from Grades VIII to XII based on national selections of from 750 to 2000 cases for each grade sex group for Form A and from 350 to 1100 cases for Form B. The manual has illustrative case studies which offer assistance to counselors in the use of results. Much research is needed to determine the validity of profiles for the prediction of various kinds of educational and vocational success.

Guilford and Zimmerman (22) constructed the *Guilford-Zimmerman Aptitude Survey* which, at present, consists of seven factors. These authors state that when the survey is complete it will include approximately twenty primary abilities. At this time, however, the seven available factors are verbal comprehension, general reasoning, perceptual speed, ability to do rapid and accurate work with numbers, spatial relations, spatial visualization, and mechanical experience. The reliabilities of these tests, determined by equivalent halves and corrected with the Spearman-Brown

formula, range from .38 to .92. As yet no estimates of validity are available. Employing the technics of factor analysis on the Army Air Forces data, Guilford and Zimmerman (23) reported the isolation of eleven intellectual factors, eight perceptual factors, three psychomotor factors, three informational factors, and two miscellaneous factors.

Dvorak (17) described the new *USES General Aptitude Test Battery*, which contains fifteen tests: eleven paper-and-pencil and four apparatus tests. The tests are intelligence (G), verbal ability (V), numerical ability (N), spatial abilities (S), form perception (P), clerical perception (Q), aiming (A), motor speed (T), finger dexterity (F), and manual dexterity (M). These tests were correlated with production records as a criterion of occupational success. Norms are available for twenty fields of work representing about 2000 occupations. These norms consist of minimum aptitude scores required for occupations and are expressed as occupational aptitude patterns. An individual aptitude profile is obtained, which may be compared with the available occupational patterns. Limitations of the tests are the lack of artistic, musical, and eye-hand-foot coordination tests.

In addition to the three differential aptitude test batteries described above, several multiple aptitude test batteries have been published or revised. Cleeton and Mason (9) revised the *Vocational Aptitude Examination, Type E-A*, designed to measure aptitudes and interests basic to the development of executive, sales, accounting, and technical skills. This battery is intended for industrial use and for the guidance of college students and adults. The first six tests measure general information, arithmetical reasoning, judgment in estimating, symbolic relationships, reading comprehension, and vocabulary. Test 7 is a short interest inventory and Test 8 is an eighty-item questionnaire on social responsiveness. No total score is obtained but a profile may be drawn.

The *Roeder General Aptitude Profile* (37) is a multiple aptitude test battery in the clerical, computational, mechanical, social service, scientific, and persuasive fields. A profile is obtained and may be compared with a list of occupational patterns.

Segel (40) reported validity data on the secondary-school level for a multiple aptitude test battery which was adapted from aptitude tests developed in the War Department. The data obtained were based on the administration of the adapted tests to representative school populations. The battery contains seven tests: mechanical aptitude, spatial relationship, speed of perception, code learning, word fluency, language usage, and mathematical reasoning. Correlational data showing the relationship between subtest scores and success in various school subjects were presented.

By means of the technic of factor analysis, Diamond (13) showed that the *Wechsler-Bellevue Intelligence Scale* can be useful as a multiple or vocational aptitude test. The linguistic factor includes the total weighted score of the Information, Comprehension, and Similarities subtests. The

clerical factor includes the total weighted score of the Digit Span, Arithmetic, and Digit Symbol subtests. Likewise, the spatial factor includes the Picture Completion, Object Assembly, and Block Design subtests. Scores in these factors were correlated with the *O'Rourke Survey Test for Vocabulary*, the *Minnesota Clerical Test*, and the *Minnesota Spatial Relations Test*. There was shown to be a fairly high degree of correspondence between each of the groups of *Wechsler-Bellevue* subtests and the aptitude test with which it was correlated.

Hayes (24) devised the *Vocational Aptitude Tests for the Blind*, which includes mazes and form boards, a musical aptitude test, a mechanical aptitude test, and a scholastic aptitude test.

Subjectmatter Aptitude Tests

The number of tests constructed during this period to predict success in specific subjectmatter fields is relatively small. Brueckner (7) published an arithmetic readiness test and correlated scores in the test with performance on an achievement test administered at the end of the year to pupils in Grades I and II. Baldwin (3) developed an *Inductive Reasoning Test* and correlated the test scores of pupils with their ranks in high-school mathematics. Nardi (31) designed a test to predict success in the study of the Hebrew language. Grime (21) reported in a study of the *Iowa Algebra Aptitude Test* a correlation of .68 between scores in the test and grades in first semester algebra. Krathwohl (26) reported satisfactory experience with the *Iowa Mathematics Aptitude Test* in the prediction of success in engineering.

In the area of technical education Drew (15) in a study of 559 eleven-year-old boys concluded that selection of boys for technical education can be made at age eleven by measuring technical aptitude in addition to general and verbal ability. Bradford (6) concluded that performance tests give a poor indication of success in technical school.

The scarcity of new subjectmatter aptitude tests probably reflects the growing trend toward the use of measures of previous achievement as predictors of success. Various research studies have shown that scores on educational achievement tests and grades in related fields of study offer a sound means for predicting success in various subjectmatter fields. Aptitude tests in science, foreign languages, mathematics, and vocational subjects have performed very well as predictive measures, but they have not offered some of the advantages of differential aptitude test batteries and achievement test batteries. While the latter types of tests may give somewhat lower correlations with criterion measures, they probably offer more useful information per unit of time than do subjectmatter aptitude tests. It seems, therefore, that there will be a continuing decline in the use of the subjectmatter aptitude type of test and a greater emphasis upon differential aptitude tests and achievement test batteries.

Mechanical Aptitude and Dexterity Tests

The category of mechanical aptitude and dexterity tests continues to include a wide variety of instruments. The tests may be roughly grouped into types which measure mechanical knowledge and tool information; basic skills, such as ability to read directions and to solve arithmetic problems; spatial relations thinking; manual dexterity; and mechanical comprehension. In addition to these groups of tests there are, of course, several tests which measure various combinations of the above abilities.

Among the new tests which measure primarily mechanical knowledge and tool information are those of Lawshe and others (27) and Schwalm (38). Cottingham (11) found the *Stenquist Mechanical Aptitude Tests* and *Detroit Mechanical Aptitudes Examination* to be a suitable combination for the prediction of scores in a woodworking performance test.

A new test of spatial thinking, the *Miami-Oxford Curve-Block Series*, was described in a preliminary report by Mellenbruch (30). The test contains six rectangular blocks cut lengthwise to form a number of irregular sections, resembling the *O'Connor Wiggly Block Worksample*. The score is determined by reference to the time required to complete the series and the number of errors made. The *Minnesota Paper Form Board* was revised by Likert and Quasha (28), the directions and scoring having been simplified and the practice problems having been added. This test probably remains the most widely used test of spatial relations thinking.

In the group of manual dexterity tests the *Purdue Pegboard* was further validated and standardized by Tiffin and Asher (48). Extensive norms on several male and female populations in industry were provided and the scores were correlated with performance in various industrial jobs requiring manipulation dexterity. Additional data on the Purdue test were reported by Strange and Sartain (42). Tuckman (50) reported that the performance of students on the *Minnesota Rate of Manipulation Test* was improved by testing students in groups of two. The reliability was not increased. Seashore (39) found that scores on the *Minnesota Rate of Manipulation Test* were unrelated to scholastic aptitude and that college men made a higher average score on the test than did the normative population, possibly because of their generally higher maturation.

Two tests which measure functions somewhat more comprehensively than do the strictly dexterity tests are worthy of note. The *Van Der Lugt Scale for the Measurement of Manual Ability* (51) tests the factors of speed, pressure, accuracy, motor memory, and static and dynamic coordination. The test was first standardized in Holland, but American norms have been provided. The test results showed an increase in scores with increasing age, at least until age twelve. Correlations with various other functions were also presented. The *Oseretsky Tests of Motor Proficiency* were also used to measure motor maturation (32). These tests, first published in Russia

in 1923, measure general static coordination, dynamic coordination of hands, general dynamic coordination, motor speed, simultaneous voluntary movements, and synkinesis (associated involuntary movements).

Included among the tests measuring several "factors" associated with mechanical aptitude are the *Prognostic Test of Mechanical Abilities* by Wrightstone and O'Toole (54) and the *SRA Mechanical Aptitude Test* (36). The former test, designed for Grades VII to XII, has five parts: arithmetic computation, reading drawings and blueprints, identification and use of tools, spatial relationships, and checking measurements. The latter test includes problems on tool usage, space visualization, and shop arithmetic.

The new mechanical aptitude tests and the published research of the period do not reveal anything startlingly new in the field. The newer tests reflect the modernization which has taken place in testing generally, but they do not represent distinctly new ideas comparable to those in the early tests by Stenquist. It is fairly clear that what we call mechanical aptitude is rather complex and consists of a number of factors, some mental and some motor. Developments in the field are likely, therefore, to be influenced primarily by present research in primary abilities on the one hand and motor coordination on the other. Perhaps combinations of tests comprising differential aptitude test batteries will constitute the mechanical aptitude tests of the future.

Clerical Aptitude Tests

The field of clerical aptitude tests has been characterized primarily by the study of the complex of abilities which make up what is called clerical aptitude. Baldwin (2) designed a *Clerical Perception Test* for Grades IX thru XII. Test 1 includes two number-checking parts of 100 items each and Test 2 contains two name-checking parts of 100 items each. A level of aspiration score is obtained for Part B of each test. The test also purports to detect examinees suffering from eye-strain.

Hosler (25) made a study of the interrelationships of the *Turse Short-hand Aptitude Test*, the *ERC Stenographic Aptitude Test*, and the *Henmon-Nelson Test of Mental Ability*. He also correlated scores in the three tests with measures of stenographic proficiency. It was concluded that there is a high degree of correlation between the two aptitude tests ($r = .79$), a marked relationship between the aptitude tests and the *Henmon-Nelson* (r 's of .64 and .65), and a marked relationship between the aptitude tests and stenographic achievement (r 's of .63 and .59).

Professional Aptitude Tests

During the period 1946-1949 large numbers of students sought admission to professional schools, notably medicine, dentistry, and veterinary medicine. The result has been an increased interest in careful selection

of students. Various attempts have been made, therefore, to develop new professional aptitude tests or to revise old ones.

In 1946 the Association of American Medical Colleges discontinued the use of the *Moss Medical Aptitude Test* and substituted the *Professional Aptitude Test* and an achievement examination in biological science published by the Graduate Record Examination. In 1947 a test designed to measure the applicant's understanding of social science was added. In 1948 the Educational Testing Service assumed responsibility for administering the testing program for the Association and instituted the use of the *Medical College Admissions Test* (45). Published validation studies of the new tests have been few in number (41, 55).

The Council on Dental Education (34) announced a new program of aptitude testing in dental schools. The tests were first given in the fall of 1946 but are not being employed as a condition of admission. Data will be gathered over a five-year period and a program of selection devised which will be based upon these findings. The factors being studied are: dental reading, memorization of verbal and visual material, knowledge of both general and scientific word meanings, mental ability, visualization of patterns without drawings, oral and written expression, and hand and finger dexterity.

A study of the prediction of success in schools of pharmacy was also instituted recently (35). The usefulness in prediction of the *American Council on Education Psychological Examination*, achievement tests, the *Kuder Preference Record*, and a personal data blank are being investigated. It is hoped that construction of a selection instrument can be based upon this research.

The *Iowa Legal Aptitude Test*, constructed in 1942, was revised and issued in experimental form in December 1946. The first edition was released in 1948. Adams and Stuit (1) reported favorable results for the test as used in several schools. The Educational Testing Service published a law admissions test in 1948 which is being used on an experimental basis. Law schools which desire to do so may ask prospective students to take the test on dates announced by the Educational Testing Service.

The prediction of success in engineering was studied with a variety of instruments. Coopridge and Laslett (10) did not find the *Stanford Scientific Aptitude Test* to be superior to the *American Council on Education Psychological Examination* in predicting grades in engineering and science courses. Vaughn (52) reported favorable results with the Engineering Aptitude series of the Graduate Record Examination.

Owens (33) very recently published data concerning an aptitude test for veterinary medicine. In the research culminating in the development of the aptitude test Owens studied the predictive value of previous scholastic records, psychological tests, and parts of the *Moss Medical Aptitude Test*. The new aptitude test proved to be superior to any of the other predictive indices, a validity coefficient of .62 being reported for the test.

In other professional areas, Warren and Canfield (53) designed an *Optometric Aptitude Test* consisting of measures of mathematical ability, general intellectual ability, clerical aptitude, persistence of attention, and carefulness of work habits. Traxler (49) described a new program of aptitude and achievement testing in accounting, and Bowers (5) designed an aptitude test for elementary-school teachers. An interesting feature of the latter test is the use of a concealed interview in which the examiner rates the student on several personal qualities.

Music, Art, and Visual Aptitude Tests

In the field of music Lundin (29) developed a set of musical aptitude tests consisting of measures of interval, discrimination, melodic transposition, mode discrimination, melodic sequences, and rhythmic sequences. A correlation of .51 was reported between total scores in the tests and performance. Correlations with the *Seashore Measures of Musical Talent* were low but positive. Cox (12) designed a test to measure functions somewhat similar to the *Seashore* test but reported no validity or reliability data. Bugg and Herpel (8) found a correlation of .65 between tonal memory scores and scores in the *Oregon Musical Discrimination Test* and a correlation of .61 between tonal memory scores and the *Kwalwasser-Dykema Test of Tonal Movement*.

Dimmick (14) developed a color aptitude test to measure color matching ability. Dubois and Gleser (16) constructed new measures of spatial thinking. Graves (20) developed a *Design Judgment Test* which is based upon a knowledge and appreciation of the basic principles of "aesthetic order." The test distinguished effectively between majors in art and other students. The *Farnsworth Dichotomous Test for Color Blindness* (18) was designed to measure color blindness more effectively than the Ishihara, a similar test. It was stated that the use of the tests results in the selection of fewer false positives—persons who appear to be color blind but who actually are not.

Summary

According to Stuit (44), the three most significant developments in aptitude testing during the period covered in this issue of the REVIEW are as follows: (a) the construction and publication of differential aptitude tests, (b) the more careful construction of tests, particularly with respect to the preparation of more precisely defined norms, and (c) the greater realization of the importance of the criterion in the validation of aptitude tests. As shown in one study (43), the validity coefficients reported for a test are markedly affected by the nature of the criterion. While most of the above developments did not originate during this period, they did receive increased attention. The indications are that they will in the future continue to merit the attention of careful test research workers.

Bibliography

1. ADAMS, MICHAEL, and STUIT, DEWEY B. "The Predictive Efficiency of the 1946 Revision of the Iowa Legal Aptitude Test." *Educational and Psychological Measurement* 9: 23-28; Spring 1949.
2. BALDWIN, G. BERNARD. *Clerical Perception Test*. Minneapolis, Minnesota: Educational Test Bureau, 1947.
3. BALDWIN, G. BERNARD. *Inductive Reasoning Test*. Minneapolis, Minnesota: Educational Test Bureau, 1946.
4. BENNETT, GEORGE K.; SEASHORE, HAROLD G.; and WESMAN, ALEXANDER G. *Differential Aptitude Tests*. New York: Psychological Corporation, 1947.
5. BOWERS, HENRY. *Manual Descriptive of the Aptitude Test for Elementary School Teachers-in-Training*. Revised edition. Toronto, Ontario: J. M. Dent and Sons, 1948.
6. BRADFORD, EDOUARD J. G. "Selection for Technical Education." *British Journal of Educational Psychology*, Part II, 16: 69-81; June 1946.
7. BRUECKNER, LEO J. "The Development and Validation of an Arithmetic Readiness Test." *Journal of Educational Research* 40: 496-502; March 1947.
8. BUGG, EUGENE G., and HERPEL, LLOYD. "The Significance of Tonal Memory for Musicality." *Journal of General Psychology* 35: 3-15; July 1946.
9. CLEETON, GLEN U., and MASON, CHARLES W. *Vocational Aptitude Examination, Type E-A*. Revised edition. Bloomington, Illinois: McKnight and McKnight, 1946.
10. COOPRIDER, HAROLD A., and LASLETT, HERBERT R. "Predictive Values of the Stanford Scientific and the Engineering and Physical Science Aptitude Tests." *Educational and Psychological Measurement* 8: 683-87; Winter 1948.
11. COTTINGHAM, HAROLD F. "Paper and Pencil Tests Given to Students in Woodworking." *Occupations* 27: 95-99; November 1948.
12. COX, IAN. "Tone Deafness." *Music Educators Journal* 34: 62-63; February-March 1948.
13. DIAMOND, SOLOMON. "Wechsler-Bellevue Intelligence Scales and Certain Vocational Aptitude Tests." *Journal of Psychology* 24: 279-82; October 1947.
14. DIMMICK, FORREST L. "A Color Aptitude Test, 1940 Experimental Edition." *Journal of Applied Psychology* 30: 10-22; February 1946.
15. DREW, L. J. "An Investigation into the Measurement of Technical Ability." *Occupational Psychology* 21: 34-48; January 1947.
16. DUBOIS, PHILIP H., and GLESER, GOLDINE. "The Object-Aperture Test: A Measure Involving Visualization in Three Dimensions." *American Psychologist* 3: 363; August 1948.
17. DVORAK, BEATRICE J. "New United States Employment Service General Aptitude Test Battery." *Journal of Applied Psychology* 31: 372-76; August 1947.
18. FARNSWORTH, DANIEL. *Farnsworth Dichotomous Test for Color Blindness*. New York: Psychological Corporation, 1947.
19. FLANAGAN, JOHN C., editor. *The Aviation Psychology Program in the Army Air Forces*. AAF Aviation Psychology Program Research Reports, No. 1. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1948. 316 p.
20. GRAVES, MAITLAND. "Design Judgment Test." *Journal of Consulting Psychology* 13: 66; February 1949.
21. GRIME, HERSCHEL E. "Aptitude and Ability in Elementary Algebra." *School Science and Mathematics* 47: 781-84; December 1947.
22. GUILFORD, J. PAUL, and ZIMMERMAN, WAYNE S. "Guilford-Zimmerman Aptitude Survey." *Journal of Applied Psychology* 32: 24-34; February 1948.
23. GUILFORD, J. PAUL, and ZIMMERMAN, WAYNE S. "Some Army Air Force Findings Concerning Aptitude Factors." *Occupations* 26: 154-59; December 1947.
24. HAYES, SAMUEL P. *Vocational Aptitude Tests for the Blind*. Watertown, Massachusetts: Perkins Institution, 1946.
25. HOSLER, RUSSELL J. "Aptitude Testing in Shorthand." *Journal of Business Education* 22: 25; May 1947.
26. KRATHWOHL, WILLIAM C. "Predictions of Average Class Achievement by Means of Aptitude Tests." *Journal of Engineering Education* 37: 234-42; November 1946.

27. LAWSHE, CHARLES H., JR., and OTHERS. "Purdue Mechanical Adaptability Test." *Journal of Applied Psychology* 30: 442-53; October 1946.
28. LIKERT, RENSIS, and QUASHA, WILLIAM H. *Revised Minnesota Paper Form Board Test, 1948 Manual*. New York: Psychological Corporation, 1948.
29. LUNDIN, ROBERT W. "The Development and Validation of a Set of Musical Ability Tests." *American Psychologist* 2: 350; August 1947.
30. MELLENBURCH, PAUL L. "A Preliminary Report on the Miami Oxford Curve—Block Series." *Journal of Applied Psychology* 30: 129-34; April 1946.
31. NARDI, NOACH. "Test To Measure Aptitude in the Hebrew Language." *Journal of Educational Psychology* 38: 167-76; March 1947.
32. OSERETSKY, NICHOLAS. *Oseretsky Tests of Motor Proficiency*. Minneapolis, Minnesota: Educational Test Bureau, 1946.
33. OWENS, WILLIAM A. "Development of a Test of Aptitude for Veterinary Medicine." *American Psychologist* 4: 240; July 1949.
34. PETERSON, SHAILER A. "Aptitude Testing Program of the American Dental Association." *Colleges and Universities* 23: 212-16; January 1948.
35. REMMERS, HERMANN H., and GAGE, NATHANIEL L. "The Predictive Testing Program of the Pharmaceutical Survey." *American Journal of Pharmaceutical Education* 11: 54-62; January 1947.
36. RICHARDSON, BELLOW, HENRY, and COMPANY. *Staff. SRA Mechanical Aptitudes*. Chicago: Science Research Associates, 1948.
37. ROEDER, WESLEY L. "The Roeder General Aptitude Profile." *American Psychologist* 2: 351; August 1947.
38. SCHWAM, R. A. "Industrial-Arts Aptitude Test." *Industrial Arts and Vocational Education* 35: 321-24; September 1946.
39. SEASHORE, HAROLD G. "Superiority of College Students on the Minnesota Rate of Manipulation Test." *Journal of Applied Psychology* 31: 249-53; June 1947.
40. SEGEL, DAVID. "Validity of a Multiple Aptitude Test at the Secondary School Level." *Educational and Psychological Measurement* 7: 695-705; Winter 1947.
41. SHOEMAKER, HAROLD A., and ROHRER, JOHN H. "Relationship Between Success in the Study of Medicine and Certain Psychological and Personal Data." *Journal of the Association of American Medical Colleges* 23: 190-201; May 1948.
42. STRANGE, JACK R., and SARTAIN, AARON Q. "Veterans' Scores on the Purdue Pegboard Test." *Journal of Applied Psychology* 32: 35-40; February 1948.
43. STUIT, DEWEY B. "The Effect of the Nature of the Criterion Upon the Validity of Aptitude Tests." *Educational and Psychological Measurement* 7: 671-76; Winter 1947.
44. STUIT, DEWEY B. "Significant Trends in Aptitude Testing." *Trends in Student Personnel Work*. Williamson, Edmund J., editor. Minneapolis: University of Minnesota Press, 1949. p. 62-79.
45. STUIT, DEWEY B., and SCHLICHER, RAYMOND J. *Handbook for Advisers to Students Planning To Enter Medicine*. Chicago: Association of American Medical Colleges, 1948. 34 p.
46. STUIT, DEWEY B., editor. *Personnel Research and Test Development in the Bureau of Naval Personnel*. Princeton, New Jersey: Princeton University Press, 1947. 513 p.
47. SUPER, DONALD E. *Appraising Vocational Fitness*. New York: Harper and Brothers, 1949. 727 p.
48. TIFFIN, JOSEPH, and ASHER, ESTON J. "Purdue Pegboard: Norms and Studies of Reliability and Validity." *Journal of Applied Psychology* 32: 234-47; June 1948.
49. TRAXLER, ARTHUR E. "Objective Measurement in the Accounting Field—a Second Report." *Collegiate News and Views* 2: 4-7; October 1948.
50. TUCKMAN, JACOB. "A Comparison of the Reliability and Performance for the Minnesota Rate of Manipulation Test for Subjects Tested Individually and in Groups of Two." *Journal of Applied Psychology* 30: 37-41; February 1946.
51. VAN DER LUGT, MARIA J. A. "The Van Der Lugt Psychomotor Scale for the Measurement of Manual Ability." *American Psychologist* 3: 256; July 1948.
52. VAUGHN, KENNETH W. "The Projects in the Graduate Record Office." *National Projects in Educational Measurement*. Report of the 1946 Invitational Conference on Testing Problems. American Council on Education Studies, Series I, No. 28. Washington, D. C.: the Council, 1947. p. 41-50.

53. WARREN, NEIL D., and CANFIELD, ALBERT A. "An Optometric Aptitude Test." *Educational and Psychological Measurement* 8: 183-91; Summer 1948.
54. WRIGHTSTONE, J. WAYNE, and O'TOOLE, CHARLES E. *Prognostic Test of Mechanical Abilities, Form A*. Los Angeles: California Test Bureau, 1947.
55. YOUNG, RICHARD H., and PIERSON, GEORGE A. "The Professional Aptitude Test, 1947; a Preliminary Evaluation." *Journal of the Association of American Medical Colleges* 23: 176-79; May 1948.

CHAPTER IV

Construction and Educational Significance of Structured Inventories in Personality Measurement

ARTHUR E. TRAXLER and ROBERT JACOBS

Generalizations Concerning Personality Appraisal

COMPARISON of recent research on inventories of personality qualities with that published a decade ago indicates that lately there have been relatively few studies in which these inventories have been applied with uncritical acceptance in the study of educational problems. Many more studies directed toward the evaluation and improvement of the instruments themselves have appeared. This is a desirable trend which may eventually bring about some noteworthy improvements in technics for the appraisal of personality thru structured procedures. During the period covered by this issue of the REVIEW, much less research was published on structured inventories of personality than on unstructured or projective technics. However, several new inventories of personal qualities were issued and a considerable number of studies of existing inventories was made available.

As Strang and Pansegrouw noted in the December 1948 issue of the REVIEW, the recent literature contains comparatively few studies of the older inventories, such as the Bernreuter *Personality Inventory*, the Bell *Adjustment Inventory*, and the *Allport-Vernon Study of Values*, while certain newer inventories have been the subject of a large amount of research. During this three-year period, the *Minnesota Multiphasic Personality Inventory* was studied much more extensively than was any other structured device for personality appraisal.

A considerable proportion of the published articles in this field during the period of this REVIEW was based on research in the armed forces. Since psychological research in the armed forces was summarized in the December 1948 special edition of the REVIEW, as well as elsewhere, studies of military use of structural inventories will be omitted from the present REVIEW unless they have educational significance.

Summaries and Bibliographies

Ellis (21) summarized studies of the validity of personality questionnaires in civilian use and arrived at the general conclusion that paper-and-pencil questionnaires suitable for group administration are of doubtful value in distinguishing between groups of adjusted and maladjusted persons. Heinlein (33) questioned Ellis' assignment of verbal categories to Pearsonian correlation coefficients of different degrees of magnitude. Ellis and Conrad (22) reviewed studies of the validity of personality inven-

tories in military practice and commented on the relatively favorable results in comparison with the use of these inventories in civilian situations. Humm and Humm (39) criticized certain conclusions drawn by Ellis and Conrad relative to the limited usefulness of standardized general instruments for measuring personality and urged that the results of these instruments be interpreted on the basis of a multidimensional examination of the profile scores of individual subjects. In the January 1947 issue of the REVIEW, Ellis summarized critical reviews of personality questionnaires and reported on new and revised questionnaires issued. He indicated that, in general, studies report satisfactory reliability but low validity for these questionnaires.

New Inventories of Personal Qualities

Two noteworthy new structured inventories were published for general use during the period under review. The *Kuder Preference Record—Personal* (46) was issued in a format similar to the well-known *Kuder Preference Record—Vocational*. Kuder-Richardson reliabilities ranging from .79 to .89 were reported in the Examiner's Manual for preference scores in five areas: sociable, practical, theoretical, agreeable, and dominant. Heston (37) published his *Personal Adjustment Inventory*, which provides scores for analytical thinking, sociability, emotional stability, confidence, personal relations, and home satisfaction. The manual for this test reports reliabilities varying from .80 to .91 for the six parts. The validity data for these two new inventories are not extensive.

New revisions of structured inventories which have been in use for some years include the Runner-Seaver *Personality Analysis* (66), which yields a profile of scores for nineteen aspects of personality, and the *Cowan Adolescent Adjustment Analyzer* (13), which is scored for nine categories of maladjustment. However, there are no new published data on the reliability and validity of these instruments.

Biddle (5) described the construction of a personality questionnaire for high-school pupils covering the four areas of home background and influence, social adjustment, adjustment to school policies and work, and adjustment to teachers. He reported a Spearman-Brown reliability of .948 for the composite score on the inventory. Crown (16) described the development of a fifty-item controlled-association test which appeared to be sufficiently reliable and valid for use as a part of a battery of tests of neuroticism.

Existing Personality Inventories

New Scales and Adaptations

Several studies directed toward the adaptation and refinement of the *Minnesota Multiphasic Personality Inventory* were published during the

three-year period. On the basis of the scores of two groups of elementary psychology students, Altus (1) selected sixty items from this inventory which differentiated well between achievers and nonachievers, and, at the same time, had no apparent relationship to intelligence. Cross (15) administered a Braille edition of the inventory to fifty blind persons and found that it apparently yielded results with the blind not greatly different from those obtained with sighted persons, altho he expressed some doubt concerning the representativeness of his sample of blind persons.

A number of studies of new scales for the *MMPI* dealt with scales for the detection of malingering and deception. These will be reported later in the chapter.

Applications of Factor Analysis

Factorial analysis studies are related to the question of the need for new scales for existing inventories and the need for completely new measurement devices. Martin (55) carried on a factor analysis of the *Bernreuter Personality Inventory* by Thurstone's centroid method and found that two factors which were best represented by the Flanagan F-1 and F-2 scales accounted for most of the measuring effectiveness of the inventory. Burt (6) made a factorial study of the temperamental traits of approximately 500 normal children and 300 psychoneurotic cases and discerned at least three factors—a general factor of emotionality, a bipolar factor distinguishing extrovertive from introvertive emotions, and a bipolar factor distinguishing pleasurable and unpleasurable emotions. Cattell (8) published his book, *Description and Measurement of Personality*, based in part on his own extensive research on the factors of personality. Cattell (9) also reported the outcome of a study of primary personality factors in the realm of objective tests. He reviewed the worth of forty-eight individual tests designed for the experiment and indicated fourteen factors which it was thought covered the main structure of personality, at least so far as the twenty- to thirty-year-old group was concerned. A table of correlations between thirty-five conative traits, which had been factorized by Cattell in accordance with Thurstone's centroid method, was refactorized by Banks (3) with Burt's method of summational analysis.

New Scoring Procedures

Supplementing earlier attempts by Bennett and others to simplify the scoring of the *Bernreuter Personality Inventory*, McClelland (53) gave the results of a further application of a simplified procedure he had previously reported and gave correlations obtained between the shortened scores and the full scores. Walton (78) described a new method of scoring the *Bernreuter* inventory which involved the use of special answer sheets

and scoring stencils and reported a decided saving in time, as well as an increase in accuracy.

The need for shortening the time required to score the individual form of the *Minnesota Multiphasic Personality Inventory* has occupied the attention of various users of this instrument, including Davis (18), Mullen (60), Ferguson (24), Manson and Grayson (54), and Gulde and Roy (30).

In contrast to the interest in short-cut scoring procedures for the *MMPI*, Corsini (12) found that after careful training a skilled clerk can score four or five tests an hour with acceptable accuracy in the manner outlined by Hathaway and McKinley and concluded that for relatively small numbers of *MMPI* tests there seems to be no advantage to the use of more elaborate short-cut methods.

Reliability, Validity, and Usefulness

Several studies directed toward the appraisal of inventories of personal qualities were reported during the period. Patterson (62) studied the relationship of Bernreuter scores to measures of various background factors in a group of 100 adult parents and concluded that the use of the Bernreuter as a diagnostic or research instrument is not justified. He suggested that its use be limited to obtaining leads in clinical situations or as a basis for the construction of more valid instruments. In a study using evening students in a junior college, Faw (23) found that neurotic scores on the Bernreuter inventory varied significantly for the same individuals, depending upon the situations the individuals had in mind when they answered the questions.

Seashore (71) compared scores on the *Allport-Vernon Study of Values* obtained from students in two vocational groups at the college level with subjective predictions based on the occupational activities contemplated by the students and concluded that this instrument is a valuable adjunct to other methods of appraisal in vocational counseling. Graham (29) asked teachers in a demonstration workshop to check the scores of secondary-school students on the *California Test of Personality* against their own evaluations and observations of the pupils and indicated that in most cases the ratings by the test seemed to be correct.

Various studies of the worth of the *Minnesota Multiphasic Personality Inventory* were made available. Wiener (79), using two matched groups of veterans, compared the results of the group form with the individual form of the *MMPI* and found no statistically significant differences. Modlin (59) and Walch and Schneider (77) reported favorably on the usefulness of the *MMPI* in clinical practice. The latter authors gave seven illustrative case histories. Hampton (31) concluded from the use of the *MMPI* with 407 college women that this inventory is a useful psychometric tool for classifying and diagnosing personality disorders. Meehl (56) found that in the "blind" diagnosis of *MMPI* profiles of hospitalized cases into three major abnormal categories the discriminations achieved were much better

than chance, but the proportion of false classifications was considerable. On the other hand, Hunt and others (41) studied the efficiency of standard *MMPI* profile signs in differentiating between psychotic and non-psychotic psychiatric patients and obtained negative results. The use of the K-factor failed to improve the accuracy of the diagnosis.

Heston (36) compared the masculinity-femininity scales on the *Minnesota Multiphasic Personality Inventory*, the *DePauw Adjustment Inventory*, the *Strong Vocational Interest Blank for Men*, and the *Kuder Preference Record* and found that the *MMPI* was the most effective of the four in differentiating between male and female subjects. Burton (7), however, reported that the reliability of the masculinity-femininity scale of the *MMPI* was only .70, which is too low for individual clinical use, altho his study was limited by the small number of cases used.

Cronbach (14) described a method of validating qualitative assessments of personality which involved the setting up of a criterion by obtaining careful descriptions of the manner in which the behavior to be predicted was performed and the use of chi-square in reporting statistically how well the predictions fit the appropriate criterion.

Malingering, Circumvention, Deception, and Anonymity

A universally recognized limitation of the usual paper-and-pencil inventory of personal qualities is that the "right" responses are likely to be evident to a reasonably intelligent subject. Kimber (44) attempted to determine the extent to which college students have insight into typical items in personality inventories. The instrument used was the *California Test of Personality*, Secondary Form. The results indicated that college students differ greatly in insight and that women appear to surpass men.

Some users of personality questionnaires have thought that the frankness and truthfulness of the responses could be improved by not requiring that the questionnaires be signed. A study by Damrin (17) based on the responses of high-school girls to the *Bell Adjustment Inventory* and one by Gerberich and Mason (27) using a questionnaire submitted to college students obtained results at variance with this hypothesis. In each of these studies, the differences between the signed and unsigned responses were negligible.

Recognizing that insight into inventories of personal qualities cannot be entirely controlled, psychologists have in recent years turned their attention to the application of statistical technics to detect circumvention, deception, and malingering, and to correct obtained scores for bias. Particularly noteworthy efforts of this kind have been made in connection with the *Minnesota Multiphasic Personality Inventory*. The purposes of the F-scale and the more recent K-scale for the *MMPI* are to detect the influence on test-taking attitudes and to correct for them. The construction, purpose, and application of the K-scale were discussed by Hathaway and Meehl (32, 57).

Schmidt (68) investigated the value of the K-factor in correcting the *MMPI* scores of air force soldiers and concluded that the K-scale contributed little, if anything, to differential diagnosis. Hunt (40) found that raw scores on the F-scale minus the K-scale seemed to be the best quantitative measure for differentiating malingerers from nonmalingerers. Similarly, Gough (28) found that the *F* minus *K* difference detected eighteen of twenty-two simulated *MMPI* profiles. Cofer, Chance, and Judson (11) reported that F-scores were useful in the detection of negative malingerers and that the addition of the L- and K-scores was useful for the detection of positive malingerers.

Wiener (80) reported the construction of two scoring keys, known as *S* and *O*, for five scales of the *MMPI* and indicated their usefulness in distinguishing test-taking attitudes as well as in predicting vocational and educational success. These keys were developed at the same time that Meehl and Hathaway were developing the K-scale.

Applications of Personality Inventories

As was indicated earlier, the number of studies representing applications of personality inventories to the study of educational problems was small during the period covered in this issue of the REVIEW. Owens and Johnson (61) administered the 300-item *MMPI*, the *Minnesota Personality Scale*, and a personal checklist to groups of students designated as over-achievers, normal achievers, and under-achievers. Analysis of the results revealed, according to the authors, that it was possible to isolate certain measurable traits peculiar to under-achievers and that conspicuous among these traits is social extraversion. Chyatte (10) used the *Minnesota Multiphasic Personality Inventory* to study the personality traits of professional actors and found some justification for ascribing unusual personality patterns to this occupational group. Dodge (20), using his own occupational personality inventory, investigated the personality traits of successful high-school teachers and obtained results tending to confirm an earlier study by the same author.

In a study of the prediction of success in teaching, Seagoe (70) obtained evidence favorable to the prognostic value of certain personality inventories, including the *Humm-Wadsworth Temperament Scale*, the *Bell Adjustment Inventory*, the *Bernreuter Personality Inventory*, and the *Thurstone Personality Schedule*.

The interests and personality traits of 274 Bible institute students were studied by Kimber (45) on the basis of the *California Test of Personality*, Adult Form, the *Minnesota Personality Scale*, and the *Kuder Preference Record*. The pattern of results of this group indicated high social standards, high sense of personal worth, high interest in social service, with low freedom from nervous tendencies and low computational and clerical interests.

Measures of Interests

Reviews

The educational research carried on in Great Britain with regard to interest and attitude measurement during a fifteen-year period was reviewed by Schonell (69). Ellis and Gerberich reported research in this area in the February 1947 issue of the REVIEW.

New and Revised Interest Inventories

A *Work Preference Inventory*, designed to measure ten personality traits and eight interest areas, was published by Henderson (34). Lovell (51) compared the scores of eighty-two college students on this inventory with the ratings of counselors and concluded from the biserial correlations for nine traits that the inventory is of doubtful usefulness as a clinical tool in college counseling.

Kuder (47) made Form CH of the *Kuder Preference Record—Vocational* available. This form is similar to Forms BB and BM, but it can be scored with a scale for outdoor activities in addition to the nine scales with which the other forms are scored. The new form also yields a verification score which is intended to assist users in detecting individuals who fail to follow directions or who give careless or misleading responses.

A short form of the *Kuder Preference Record* was proposed by Miles (58). He experimented with the use of pages 7, 8, and 9 on the preference record as a basis for the prediction of scores on the entire record and concluded that this short form yielded an adequate indication of the areas in which the subject had the greatest interest and that a great saving in scoring and administration time was effected. The time required for scoring the preference record was also considered by Lauro (48), who made suggestions for punching keys for machine scoring of the *Kuder* so that thirty answer sheets could be scored in seven minutes, which in his estimation resulted in a saving of 11 percent over the usual machine-scoring time.

The Kuder Preference Record

The larger part of the published research on interest inventories during this period has been concerned with the *Kuder Preference Record*. The need for better norms on the *Kuder* for college students was indicated by Heston (35), who reported percentile norms based on 672 men and 1027 women entering the freshman class at DePauw University. Perry and Shuttleworth (63) reported differences of considerable size between results of the preference record for 669 freshmen in the College of the City of New York and the national norms, but they found good agreement between degree objectives and *Kuder* results when "expected" profile patterns were used as the criterion. Baggaley (2) analyzed the *Kuder* scores of 155 Harvard freshmen and obtained evidence that these results differentiated between students concentrating in different academic fields.

Some data on the relationship between stated vocational choice and measured interest pattern were made available. Bateman (4) studied the relationship between stated vocational choice and *Kuder* profiles for two groups of junior-high-school students defined as "work experience" and "nonwork experience" groups and found closer agreement for nonworkers than for workers. Using an unselected group of sixty veterans, Rose (65) found an over-all correlation of .61 between the ranked order of the strength of the nine interest areas on the *Kuder* and the veterans' ranking of lists of occupations corresponding to the nine areas. From this finding he concluded that occupational list selection does not reliably indicate the kind of interest activity patterns that are measured by the *Kuder*.

Several studies were reported in which the *Kuder Preference Record* was applied in the appraisal of the interests of particular groups. Two studies of the measured interests of nurses were reported by Triggs (75, 76). In a study of *Kuder* results for seniors in the Indiana University School of Business, Shaffer (72) found this inventory useful in assisting the individuals to choose a major curriculum in the field of business. DiMichael (19) found correlations ranging from .24 to .75 for *Kuder* scores and the self-estimated interests of vocational rehabilitation counselors. In a study of the use of the *Kuder Preference Record* among independent school pupils, Traxler (74) found that the median reliability of the nine scales was about .85, that the intercorrelations among the scores were comparatively low, that there were slight differences among the median scores for Grades IX thru XII, and that the provision of separate sex norms was justified by the data.

A few studies have been concerned with the stability of vocational preferences at various maturity levels. In a study of the relationship between *Kuder Preference Record* scores and factors drawn from the personnel records of high-school pupils, Jackson (42) concluded that when elections of courses are made in Grade IX the choice tends to crystallize and stabilize vocational interests at that level. Fox (25) studied the stability of *Kuder Preference Record* interests among ninth-grade pupils during an eight-week period and found a rather marked degree of stability. A similar study by Jacobs (43) dealing with pupils tested in the ninth grade and again in the tenth grade yielded fairly high correlations for test-retest scores on the nine *Kuder* scales.

The fakability of the *Strong Vocational Interest Blank* and the *Kuder Preference Record* was studied by Longstaff (50) thru an analysis of the results of these inventories administered to students in an evening class at the University of Minnesota. The results indicated that both tests were definitely fakable and that some sort of empirical scale to detect faking was needed.

Reading Difficulty of Interest Inventories

Questions are often raised by counselors and others concerning the lowest age or grade level at which various interest inventories may profit-

ably be used. Studies of the vocabulary burden and the reading difficulty of these inventories are useful in answering questions of this kind. Steffire (73) rated the Brainard, Kuder, Lee-Thorpe, Allport-Vernon, Strong, and Cleeton vocational interest inventories by the Lewerenz formula for vocabulary grade placement. The data indicated that about seventh-grade reading ability was needed to understand the Brainard, Kuder, and Lee-Thorpe inventories; that at least tenth- or eleventh-grade reading ability was required to comprehend the Allport-Vernon and the Strong blanks; and that the Cleeton inventory was at the twelfth-grade level. Roeber (64) compared the Brainard, Cleeton, Garretson-Symonds, Kuder, Lee-Thorpe, Strong, and Thurstone inventories with respect to word usage and found that at least 10 percent of the words in all these inventories were beyond the recommended range for ninth-grade vocabulary. He indicated some precautions which may be taken in administering any of these inventories at the ninth-grade level.

Inventories of Other Personal Qualities

Attitude Scales

A battery of feeling and attitude scales for clinical use was released by Hildreth (38). Scales for six feeling states and attitudes were constructed by a modified Thurstone technic. Most of the intercorrelations among the scales fell within the range of .20 to .50. Altho these scales were developed in a military hospital, Lehner and Hunt (49) administered them to 129 college students and obtained evidence "that the test may be serving somewhat the same function in both clinical and educational groups for both men and women."

The construction and application of an inventory for differentiating attitudes of high-school students were reported by Lyman (52).

Sartain and Bell (67) criticized the Bogardus *Scale of Social Distance* on the basis that it was unbalanced in number of favorable and unfavorable statements. They constructed a revised Bogardus Scale and two Thurstone-type scales, administered these four scales to 100 college students in order to measure their attitudes toward the English, Japanese, and Negroes, and reported the results.

Persistence Tests

The experimental use of a persistence test constructed at the Educational Testing Service was discussed by French (26). The purpose was to raise the multiple-correlation of a test battery with college grades used as the criterion. The new test was scored in various ways and correlated with College Entrance Test scores and course grades. The contributions of the persistence test raised the R from .58 to .65. It was felt that the results of the study pointed the way to further research.

Bibliography

1. ALTUS, WILLIAM D. "A College Achiever and Non-Achiever Scale for the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology* 32: 385-97; August 1948.
2. BAGGALEY, ANDREW R. "The Relation Between Scores Obtained by Harvard Freshmen on the Kuder Preference Record and Their Fields of Concentration." *Journal of Educational Psychology* 38: 421-27; November 1947.
3. BANKS, CHARLOTTE. "Primary Personality Factors in Women: A Re-Analysis." *British Journal of Psychology, Statistical Section* 1, Part 3: 204-18; November 1948.
4. BATEMAN, RICHARD M. "Effect of Work Experience on High School Students' Vocational Choice as Revealed by the Kuder Preference Record." *Occupations* 27: 453-56; April 1949.
5. BIDDLE, RICHARD A. "Construction of a Personality Inventory." *Journal of Educational Research* 41: 366-77; January 1948.
6. BURT, CYRIL. "The Factorial Study of Temperamental Traits." *British Journal of Psychology, Statistical Section* 1, Part 3: 178-203; November 1948.
7. BURTON, ARTHUR. "The Use of the Masculinity-Femininity Scale of the Minnesota Multiphasic Personality Inventory as an Aid in the Diagnosis of Sexual Inversion." *Journal of Psychology* 24: 161-64; July 1947.
8. CATTELL, RAYMOND B. *Description and Measurement of Personality*. Yonkers, New York: World Book Company, 1946. 602 p.
9. CATTELL, RAYMOND B. "Primary Personality Factors in the Realm of Objective Tests." *Journal of Personality* 16: 459-87; June 1948.
10. CHYATTE, CONRAD. "Personality Traits of Professional Actors." *Occupations* 27: 245-50; January 1949.
11. COFER, CHARLES N.; CHANCE, JUNE E.; and JUDSON, ABE J. "A Study of Malingering on the Minnesota Multiphasic Personality Inventory." *Journal of Psychology* 27: 491-99; April 1949.
12. CORSINI, RAYMOND J. "A Time and Motion Study of Hand Scoring the Individual Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 13: 62-63; February 1949.
13. COWAN, EDWINA A.; MUELLER, WILBERT J.; and WEATHERS, EDRA. *Cowan Adolescent Adjustment Analyzer: An Instrument of Clinical Psychology*. Salina, Kansas: Cowan Research Project, 1946.
14. CRONBACH, LEE J. "Validation Design for Qualitative Studies of Personality." *Journal of Consulting Psychology* 12: 365-74; November 1948.
15. CROSS, ORRIN H. "Braille Edition of the Minnesota Multiphasic Personality Inventory for Use with the Blind." *Journal of Applied Psychology* 31: 189-98; April 1947.
16. CROWN, SIDNEY. "A Controlled Association Test as a Measure of Neuroticism." *Journal of Personality* 16: 198-208; December 1947.
17. DAMRIN, DORA E. "A Study of the Truthfulness with Which High School Girls Answer Personality Tests of the Questionnaire Type." *Journal of Educational Psychology* 38: 223-31; April 1947.
18. DAVIS, CLIFFORD E. "The Minnesota Multiphasic Personality Inventory: A New Method of Scoring and Analysis." *Journal of Clinical Psychology* 3: 298-301; July 1947.
19. DiMICHAEL, SALVATORE G. "Professed and Measured Interests of Vocational Rehabilitation Counselors." *Educational and Psychological Measurement* 9: 59-72; Spring 1949.
20. DODGE, ARTHUR F. "A Study of the Personality Traits of Successful Teachers." *Occupations* 27: 107-12; November 1948.
21. ELLIS, ALBERT. "The Validity of Personality Questionnaires." *Psychological Bulletin* 43: 385-440; September 1946.
22. ELLIS, ALBERT, and CONRAD, HERBERT S. "The Validity of Personality Inventories in Military Practice." *Psychological Bulletin* 45: 385-426; September 1948.
23. FAW, VOLNEY. "Situational Variations of Neurotic Scores Measured by the Bernreuter Inventory." *Journal of Consulting Psychology* 12: 255-58; July-August 1948.
24. FERGUSON, ROBERT G. "A Useful Adjunct to the Minnesota Personality Inventory Scoring and Analysis." *Journal of Clinical Psychology* 2: 248-53; July 1946.

25. FOX, WILLIAM H. "Stability of Measured Interests." *Journal of Educational Research* 41: 305-10; December 1947.
26. FRENCH, JOHN W. "The Validity of a Persistence Test." *Psychometrika* 13: 271-77; December 1948.
27. GERBERICH, JOHN B., and MASON, JOHN M. "Signed versus Unsigned Questionnaire." *Journal of Educational Research* 42: 122-26; October 1948.
28. GOUGH, HARRISON G. "Simulated Patterns on the Minnesota Multiphasic Personality Inventory." *Journal of Abnormal and Social Psychology* 42: 215-25; April 1947.
29. GRAHAM, ALVA WHITCOMB. "Personal and Social Adjustment of High School Students." *School Review* 55: 468-73; October 1947.
30. GULDE, CARL J., and ROY, HOWARD L. "A Note on the Scoring of the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 11: 221-22; July-August 1947.
31. HAMPTON, PETER J. "The Minnesota Multiphasic Personality Inventory as a Psychometric Tool for Diagnosing Personality Disorders Among College Students." *Journal of Social Psychology* 26: 99-108; August 1947.
32. HATHAWAY, STARKE R., and MEEHL, PAUL E. *The K-Scale for the Minnesota Multiphasic Personality Inventory*. New York: Psychological Corporation, 1947. 8 p.
33. HEINLEIN, CHRISTIAN PAUL. "Comment on 'The Validity of Personality Questionnaires.'" *Psychological Bulletin* 44: 80-82; January 1947.
34. HENDERSON, ROBERT W. *Work Preference Inventory*. Lexington (University Personnel Office, University of Kentucky): the Author, 1946.
35. HESTON, JOSEPH C. "College Freshman Norms for the Kuder Preference Record." *Occupations* 26: 92-94; November 1947.
36. HESTON, JOSEPH C. "A Comparison of Four Masculinity-Femininity Scales." *Educational and Psychological Measurement* 8: 375-87; Autumn 1948.
37. HESTON, JOSEPH C. *Heston Personal Adjustment Inventory*. Yonkers, New York: World Book Company, 1949.
38. HILDRETH, HAROLD M. "A Battery of Feeling and Attitude Scales for Clinical Use." *Journal of Clinical Psychology* 2: 214-21; July 1946.
39. HUMM, DONCASTER G., and HUMM, KATHRYN A. "Notes on 'The Validity of Personality Inventories in Military Practice,' by Ellis and Conrad." *Psychological Bulletin* 46: 303-306; July 1949.
40. HUNT, HOWARD F. "Effect of Deliberate Deception on MMPI Performance." *Journal of Consulting Psychology* 12: 396-402; November 1948.
41. HUNT, HOWARD F., and OTHERS. "A Study of the Differential Diagnostic Efficiency of the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 12: 331-36; September-October 1948.
42. JACKSON, JOSEPH. "Note on the Crystallization of Vocational Interests." *Journal of Social Psychology* 26: 125-30; August 1947.
43. JACOBS, ROBERT. "Stability of Interests at the Secondary School Level." 1949 *Achievement Testing Program in Independent Schools and Supplementary Studies*. Educational Records Bulletin No. 52. New York: Educational Records Bureau, July 1949. p. 83-87.
44. KIMBER, J. A. MORRIS. "The Insight of College Students into the Items on a Personality Test." *Educational and Psychological Measurement* 7: 411-20; Autumn 1947.
45. KIMBER, J. A. MORRIS. "Interests and Personality Traits of Bible Institute Students." *Journal of Social Psychology* 26: 225-33; November 1947.
46. KUDER, G. FREDERIC. *Kuder Preference Record—Personal*. Chicago: Science Research Associates, 1948. (not paged)
47. KUDER, G. FREDERIC. *Kuder Preference Record—Vocational, Form CH*. Chicago: Science Research Associates, 1948. (not paged)
48. LAURO, LOUIS. "A Note on Machine Scoring the Kuder Preference Record." *Journal of Applied Psychology* 32: 629-30; December 1948.
49. LEHNER, GEORGE F. J., and HUNT, EDWARD L. "The Use of the Hildreth Feeling and Attitude Scales with College Students." *Journal of Psychology* 26: 217-22; July 1948.
50. LONGSTAFF, HOWARD P. "Fakability of the Strong Interest Blank and the Kuder Preference Record." *Journal of Applied Psychology* 32: 360-69; August 1948.
51. LOVELL, GEORGE D. "Validating Studies of the Work Preference Inventory." *Journal of Applied Psychology* 32: 195-99; April 1948.

52. LYMAN, HOWARD B. "Differentiating Attitudes of Students at Two High Schools by Use of a School Attitude Inventory." *Educational and Psychological Measurement* 9: 227-32; Summer 1949.
53. MCCLELLAND, DAVID C. "Further Application of Simplified Scoring of the Bernreuter Personality Inventory." *Journal of Applied Psychology* 31: 182-88; April 1947.
54. MANSON, MORSE P., and GRAYSON, HARRY M. "Keysort Method of Scoring the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology* 30: 509-16; October 1946.
55. MARTIN, GLENN C. "A Factor Analysis of the Bernreuter Personality Inventory." *Educational and Psychological Measurement* 8: 85-92; Spring 1948.
56. MEEHL, PAUL E. "Profile Analysis of the Minnesota Multiphasic Personality Inventory in Differential Diagnosis." *Journal of Applied Psychology* 30: 517-24; October 1946.
57. MEEHL, PAUL E., and HATHAWAY, STARKE R. "The K-Factor as a Suppressor Variable in the Minnesota Multiphasic Personality Inventory." *Journal of Applied Psychology* 30: 525-64; October 1946.
58. MILES, RAY W. "A Proposed Short Form of the Kuder Preference Record." *Journal of Applied Psychology* 32: 282-85; June 1948.
59. MODLIN, HERBERT C. "A Study of the Minnesota Multiphasic Personality Inventory in Clinical Practice." *American Journal of Psychiatry* 103: 758-69; May 1947.
60. MULLEN, FRANCES A. "The Minnesota Multiphasic Personality Inventory: An Extension of the Davis Scoring Method." *Journal of Clinical Psychology* 4: 86-88; January 1948.
61. OWENS, WILLIAM A., and JOHNSON, WILMA C. "Some Measured Personality Traits of Collegiate Underachievers." *Journal of Educational Psychology* 40: 41-46; January 1949.
62. PATTERSON, CECIL H. "The Relationship of Bernreuter Scores to Parent Behavior, Urban-Rural Residence, and Other Background Factors in 100 Normal Adult Parents." *Journal of Social Psychology* 24: 3-49; August 1946.
63. PERRY, JAMES D., and SHUTTLEWORTH, FRANK K. "Kuder Profiles of College Freshmen by Degree Objectives." *Journal of Educational Research* 41: 363-65; January 1948.
64. ROEBER, EDWARD C. "Comparison of Seven Interest Inventories with Respect to Word Usage." *Journal of Educational Research* 42: 8-17; September 1948.
65. ROSE, WALLACE. "A Comparison of Relative Interest in Occupational Groupings and Activity Interests as Measured by the Kuder Preference Record." *Occupations* 26: 302-307; February 1948.
66. RUNNER, JESSIE R., and SEAVER, MARGARET A. *Personality Analysis*. Iowa City, Iowa (214 Hutchinson): Jessie R. Runner, 1948.
67. SARTAIN, AARON Q., and BELL, HAROLD V., JR., "An Evaluation of the Bogardus Scale of Social Distance by the Method of Equal-Appearing Intervals." *Journal of Social Psychology* 29: 85-91; February 1949.
68. SCHMIDT, HERMANN O. "Notes on the Minnesota Multiphasic Personality Inventory: The K-Factor." *Journal of Consulting Psychology* 12: 337-42; September-October 1948.
69. SCHONELL, FRED J. "Measurement of Attitudes and Interests." *British Journal of Educational Psychology* 18: 53-61; June 1948.
70. SEAGOE, MAY V. "Prediction of In-Service Success in Teaching." *Journal of Educational Research* 39: 658-63; May 1946.
71. SEASHORE, HAROLD G. "Validation of the Study of Values for Two Vocational Groups at the College Level." *Educational and Psychological Measurement* 7: 757-63; Winter 1947.
72. SHAFFER, ROBERT H. "Measured Interests of Business School Seniors." *Occupations* 27: 462-65; April 1949.
73. STEFFLRE, BUFORD L. "The Reading Difficulty of Interest Inventories." *Occupations* 26: 95-96; November 1947.
74. TRAXLER, ARTHUR E. "Some Results of the Kuder Preference Record in Independent Schools." 1945 *Fall Testing Program in Independent Schools and Supplementary Studies*. Educational Records Bulletin No. 44. New York: Educational Records Bureau, January 1946. p. 37-50.
75. TRIGGS, FRANCES ORALIND. "The Measured Interests of Nurses." *Journal of Educational Research* 41: 25-34; September 1947.

76. TRIGGS, FRANCES ORALIND. "The Measured Interests of Nurses: A Second Report." *Journal of Educational Research* 42: 113-21; October 1948.
77. WALCH, ALPHONSE E., and SCHNEIDER, ROBERT A. "The Minnesota Multiphasic Personality Inventory: An Evaluation of Its Use in Private Practice." *Minnesota Medicine* 30: 753-58; July 1947.
78. WALTON, WILLIAM E. "A New Method of Scoring the Bernreuter Personality Inventory." *Personnel* 24: 200; November 1947.
79. WIENER, DANIEL N. "Differences Between the Individual and Group Forms of the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 11: 104-106; March-April 1947.
80. WIENER, DANIEL N. "Subtle and Obvious Keys for the Minnesota Multiphasic Personality Inventory." *Journal of Consulting Psychology* 12: 164-70; May-June 1948.

CHAPTER V

Development and Educational Significance of Projective Technics in Personality Measurement

PERCIVAL M. SYMONDS AND MARTHA G. HESSEL

THE number of articles on projective technics increased considerably in the three-year period since the last REVIEW on this topic. It was, therefore, necessary to limit this survey of the research done to studies of particular relevance to workers in education and related fields.

General Approaches to Projective Technics

A number of books and articles describing and illustrating projective technics appeared during the three-year interval, indicating the increasing interest and use of these procedures. The rationale of these methods was especially emphasized in Frank's monograph (37). He discussed the dynamic theory of personality underlying the use of these tests, the current trends in scientific thought which made their development possible, and different theoretical approaches to the diagnosis of personality. He briefly described the main technics used, classifying them as constitutive, constructive, interpretive, cathartic, and refractive. Another useful work for the student in this field was written by Bell (8). He made a comprehensive summary of the literature of many projective methods and further attempted to provide an introductory manual for many of them. The publication of Buros' yearbook (15) provided another valuable source of references. Rosenzweig and Kogan (67) gave many illustrations of the scoring and interpretation of some of the more important tests in their book on psychological testing.

The multiple use of projective technics continued. The *Rorschach*, the *Thematic Apperception Test*, and play technics were used by Merrill (54) in her study of juvenile delinquents. Havighurst and Taba (44) used the *Rorschach* and a *TAT* type test in one phase of their study of adolescent character and personality. Carp (16), rating "constriction" in the *Rorschachs*, play productions, and crayon drawings of ninety-six third-grade students, found only a chance relationship between constriction ratings for an individual child in these three factors. Bell (7), however, found that the personality picture evoked in a play situation and a picture-story situation could be matched successfully. Goodenough (38) emphasized the need for caution in the interpretation of children's fantasy products and the need for more thorough investigation of the individual child. Schafer and Leitch (71) devised a tentative list of signs for the *TAT*, the *Rorschach*, and the *Stanford-Binet Scale*, Form L, for detecting mal-

adjustment tendencies in twenty-two nursery-school children. Much more experimentation in the use of signs for differential diagnosis is definitely indicated.

An important effort was made to progress from diagnosis to prediction. In the selection of clinical psychology trainees, technics involving a number of projective methods as well as situational and real-life situations were described by Kelly (48). This selection program was modified from the procedures used during the war by the Office of Strategic Services (84). Kelly stated that this program was primarily for the evaluation of the technics for future use after they had been validated against the actual accomplishments of the trainees.

The Rorschach

The amount of research, comments, and clinical reports dealing with the *Rorschach* is truly impressive. Altho Rotter (68) stated, with others, that relatively little was being produced in terms of new validation data, the test continued to be the most popular projective technic.

Ford (36), applying the *Rorschach* to a selected group of 123 children three to eight years of age, was particularly interested in the statistical and nonclinical aspects of the method. She attempted to devise directions and a method of administration applicable to young children; to study variations in test determinants along with variations in chronological age, mental age, and sex; to study the reliability of the various test determinants at the preschool level; and to check the validity of certain claims about the meaning of various test determinants against objective criteria. Of special interest was her statistical evidence that many of the determinants on which the total personality pattern depended were closely related to the age of the subject. Test-retest reliability for fifty-five preschool children in the major determinants ranged from .38 to .86, with the majority above .70.

The multiple-choice *Rorschach* was still used during this three-year period. Lawshe and Forster (49) found the reliability of the test to be low. Engle (32) tried to differentiate, with limited success, between well-adjusted and maladjusted high-school pupils. Blair and Clark (11) attempted to estimate the personality maladjustments of 382 ninth-grade pupils. The indifferent results in these three studies tend to be in harmony with much of the previous research done on this instrument.

Rorschach patterns related to the sociometric status of forty-five eighth-grade school children were investigated by Northway and Wigdor (61). Groups differing in social status showed significant differences in the *Rorschach*. Siegel (76) discussed the diagnostic and prognostic validity of the *Rorschach* in a child guidance clinic. In the area of behavior problems, Schmidl (72) reviewed the use of the test in juvenile delinquency research and made recommendations for further research.

Jolles (46), who studied sixty-six children seriously retarded in mental development as judged by psychometric examinations, stressed the importance of personality adjustment (measured by the *Rorschach*) as a factor in the etiology of mental deficiency. More work is needed in this area to validate his results. Sarason and Sarason (70) used the test pattern of the *Rorschach*, the *Kohs Blocks*, and the *Binet* in discriminating between a group of defective cerebral palsy children and a group of familial defectives.

Partly because college populations were readily accessible to research workers, considerable work was done at the college level with the *Rorschach*. Because of extensive use of the group *Rorschach*, Munroe's discussion of the application and implications (58) of this adaptation of the test was of interest. Dunkel (28) used the group *Rorschach* in an attempt to account for the discrepancy between verbal intelligence as measured by the *American Council on Education Psychological Examinations* and linguistic performance on a placement test in Latin for entering college freshmen. Munroe (57) investigated the *Rorschach* patterns for two groups of college students, one of which had received a high score on linguistic subtests of the A.C.E., in relation to their score on the quantitative subtests, the other of which was relatively more proficient with the quantitative materials. Montalto (55) studied the relationship between academic standing and *Rorschach* signs of adjustment in ninety women college students. Using partial correlations she devised a pattern of signs on the group *Rorschach* which she felt could be used for prediction of academic achievement. Thompson (81), administering the group *Rorschach* to 128 college students, obtained a correlation of .38 between semester grades in a course in psychology and her quantified method of scoring the test. McCandless (50) found no statistical differences between thirteen pairs of men who differed in terms of academic grade average, altho trends were discovered. A series of three studies by Altus (3), Clark (20), and Thompson (82) attempted to find the relation between signs obtained from group *Rorschach* and various scales of the *Minnesota Multiphasic Personality Inventory* on a population of college students.

The *Rorschach* was used in industry, vocational studies, and evaluation of counseling. Steiner (78) reviewed five studies dealing with the use of the *Rorschach* and other projective methods in industry. The results were far from conclusive. Kaback (47) made a primarily statistical study of the vocational personalities of seventy-five men engaged in the profession of pharmacy, seventy-five in accountancy, and seventy-five students preparing for each of these areas. Using twenty-four *Rorschach* components she found the test of negligible value for predictive purposes. Anderson and Munroe (4) investigated by means of the group *Rorschach*, scored by the inspection technic, the personality factors in students interested in creative painting as compared with those interested in commercial art. Use of the *Rorschach* was made by Muench (56), Carr (17), and Hamlin and Albee (42) in evaluating the results of nondirective therapy.

An attempt was made by Gustav (41) to determine the relationship between scoring categories of the individual form of the *Rorschach* and items in various standard personality scales and then to construct an objective personality inventory based on the *Rorschach*. It was felt that the objective inventory would serve as a screening device for college freshmen.

Thematic Apperception Test and Other Picture-Projective Tests

Considerable work was done with the *TAT*, altho much of it dealt with topics outside the scope of this paper. A number of articles and books were published which presented different systems of analysis and interpretation. Stein's manual (77) described the pictures and stories most often told to them by adult males. A detailed account of the interpretation of a complete protocol was given. Tomkins in his manual (83) reviewed the literature and theory of the test and gave numerous illustrations of his method of interpretation. Wyatt (87) reviewed the types of approaches to interpretation of fantasy material and presented his own system.

One of the chief complaints leveled against the test continued to be its lack of objectivity—specifically the difficulties involved in quantifying the scoring into meaningful categories, the lack of norms, and the question of its validity. A number of studies were done at the college level in an effort to devise quantified methods of scoring, to determine the reliability of the scoring, and to establish some type of normative system. Wittenborn (85) classified the responses and the frequency with which certain content was found among 100 college men, sixteen to twenty-five years of age. He found several patterns of adjustment with vocational and educational significance. The stories told by 250 college students in response to viewing ten *TAT* pictures were analyzed by Bellak, Ekstein, and Braverman (9) to investigate the nature and frequency of principal themes. Eron (33) made a study of the frequencies of themes and identification in the stories of twenty-five schizophrenic patients and twenty-five nonhospitalized college students. The two groups differed at or beyond the 5 percent level of significance in only thirteen out of ninety-eight themes. He concluded that the thematic material obtained on the *TAT* was very much a function of the stimulus properties of the cards themselves. This study seems to indicate that the content analysis of the *TAT* has little value for differential diagnosis.

Some attempt was made to evaluate the test's validity. Combs (24), using *TAT* protocols and the autobiographies of forty-six college students, investigated the use of personal experience in *TAT* story plots. In comparing the themes he found that the subjects drew on their life experience to some extent for fantasy plots. In a study of motivation (23), he obtained a rank order correlation of .74 between the forty most common desires expressed on the *TAT* and in the autobiographies. In work with children

of eight to fourteen, Coleman (21), giving one series of *TAT* pictures before a film (adjudged "recent experience") and one series after, found that *TAT* responses were not affected by films of average dramatic intensity. Presumably *TAT* responses would not be affected very much by life events of average intensity and would therefore have some claim to stability.

Other studies using the *TAT* were not so interested in quantification and normative data. Gothberg (39) compared the *TAT* stories of ten runaway girls with a control group. Bettelheim (10), working with college students, showed how the *TAT* could be used as an educational and therapeutic device in a classroom situation by having the students analyze their own protocols.

In the past three-year period, technics and principles of the *TAT* have been adapted to investigate specific problems of interest to educators. Symonds (79, 80), in an investigation of adolescent personality, developed a set of twenty pictures designed especially to evoke adolescent fantasy. His subjects were forty normal adolescent boys and girls. Intensive case studies were made of these children so that comparisons of the child's fantasy productions and his everyday life at home and in school could be made. In his book Symonds reviewed the literature on this type of test and described the development and selection of the pictures used. He made an empirical analysis of the themes present in the stories and developed norms showing the relative frequency of the principal themes associated with the stories. Correlation methods applied to quantitative data confirmed the observation that the fantasy material was often in complete contrast to the case material and life history. Hypotheses for these discrepancies were presented in light of the objective data. The significance and characteristics of adolescent fantasy were discussed. The case of a maladjusted boy and that of a well-adjusted boy were presented with an analysis and interpretation of their stories. Eiserer (29) in a primarily methodological study investigated the relative effectiveness of motion and equivalent still pictures in eliciting fantasy stories from fifty adolescent boys about adolescent-parent relationships. He found in his analysis of the stories that school figured prominently in the lives of these adolescents. Further contributions to methodology and theory were made by McClelland and others (51) in a study of 200 male college students writing stories to four pictures after different success or failure conditions.

Blum (12), using twelve cartoon drawings showing the adventures of a dog named Blacky, attempted to study the psychoanalytical theory of psychosexual development in college students. Dorkey and Amen (26) studied anxiety in twenty-four nursery-school children by means of equivocal pictures. The manual for Rosenzweig's *Picture-Frustration Study for Adults* (65) was revised and a comparable form for children from four to thirteen years developed (66). Age norms for the various scoring categories were given for the latter form. Joel (45) employed the *Make-A-Picture Story (MAPS) Test* with fifty disturbed adolescents. In this test the subject chose figures to populate a given background and then told a story of the situation

depicted. Norms were needed to make evaluation more valid. A somewhat similar technic, used by Chein and Evans (19) and called the movie-story game, was used to investigate the interracial attitudes of Negro and white children.

Play Technics

A great deal of worthwhile research was published in this area, particularly in the investigation of methodology and quantification of the scoring. Sears (73) discussed the importance of methodology and reviewed a number of studies in this area. It was felt to be essential to know the type of reactions evoked by different methods in doll play before diagnosis of any validity could be made. Harms (43) presented another view of play analysis, altho he emphasized its relation to the child's development.

Radke (63), using a projective picture technic and doll play, among other more objective methods, investigated the relation of parental authority to the behavior and attitudes of forty-three nursery and kindergarten children. An investigation of group data revealed slight relationships between doll-play data and other data on home relationships. Robinson (64), studying the effect of the doll family constellation on doll play, made further contributions to the study of methodology. Bach (5), applying a quantified scoring system to the doll play of father-separated children and a control group, found eleven statistically significant differences between the father fantasies of the two groups. Sears, Pintler, and Sears (74) found in an investigation of doll play that children from father-absent homes did not show in the second play interview the customary rise in aggression characteristic of the control group in this study and of groups in other studies. Bach and Bremer (6), using a similar doll-play situation with twenty normally adjusted children and twelve preadolescent delinquent children, contrasted the father fantasies of these groups. The deviants showed almost indifference to the father figure. Meister (53) attempted to diagnose the adjustment of thirty-two children ranging from five to seven years in age by means of play performance with a standard doll-play situation. Further work in this area, using different populations, different technics, and different experimental conditions, would seem to be indicated.

Drawing and Painting Technics

Altho attempts were made to verify hypotheses experimentally and to set up developmental norms for children in drawing and painting, much of the work relating to personality continued to be predominantly subjective in nature. Wolff (86) made analyses of preschool children's drawings as a method of studying their personal ties. Altho many of his hypotheses were stimulating, experimental verification of group or individual data was not evident. On the basis of observation of thirty children, Elkish (30) discussed a "scribbling game" to be used in the diagnosis of children. England

(31) studied the fears of four groups of children—normal, retarded, institutional, and sex delinquent—by means of drawings. Dubin (27) found that training sped up the process of graphic representation in twenty-six preschool children. Further work in developmental factors and the effect of training should serve to make the field more objective and less subject to superficial interpretations.

An important contribution was made by Alschuler and Hattwick (2) on the relationship between painting and personality in 150 children from two to four years of age. Observation, information about their homes, and life experiences were gathered for validating data. The authors found that the children tended to express similar feelings and conflicts comparably. In Volume I the two authors discussed color choices, the use of line and form, and space usage and spatial patterns in terms of the child's dynamics expressed in these factors. They emphasized the importance of a genetic, developmental approach to children's art work. Altho particularly interested in easel painting, they compared the child's use of paints with his use of other media. Volume II presented brief biographical summaries of each child as well as quantitative findings with interpretive summaries of these. Naumberg (60) stressed the value of free art expression of behavior-problem children in diagnosis and therapy. Whereas other investigators have stressed formal factors in painting, Naumberg concentrated her attention on the significance of content.

Finger painting continued to be used as a medium for the study of personality. Napoli (59) published a finger-painting record form. Phillips and Stromberg (62) made a study of the finger-painting performance in twenty-five detention-home and twenty-five high-school pupils. Using the chi-square technic, they found that the two groups differed significantly at the 5 percent level in fourteen out of seventeen scoring categories. Blum and Dragositz (13) studied the developmental aspects of finger painting in twenty-four first-grade and twenty-nine sixth-grade pupils. Definite changes with age were found.

Miscellaneous Approaches to Personality Evaluation

In addition to the aforementioned projective technics employed in the study of personality, various other methods were of significance to workers in education. Rotter and Wickens (69) utilized role-playing situations to study the "consistency and generality of social aggressiveness." Reliabilities were found to be high enough to justify further studies. Bronfenbrenner and Newcomb (14) discussed the technic of improvisation, which was an adaptation of psychodrama specifically for personality diagnosis. They felt it to be most useful as a projective technic. Improvisations, based on standard situations, were used in the evaluation of clinical psychology candidates in the Veterans Administration. Important features of the technic were discussed, as well as the situations used and the analysis and interpretation of results.

Costin and Eiserer (25) gave a forty-item sentence completion test to seventy-four eleventh-grade pupils. The responses were classified according to the attitudes the students expressed toward school in general, specific aspects of school life, teachers, and students. More negative attitudes were expressed toward teachers than toward any other factor of school life. Grover (40) stated that teachers could gain understanding of students' personalities from their compositions, apparently by utilizing technics of projective interpretation.

Alberman and Schaeffer (1) devised four stories, specifically focused on problems of children from seven to eleven years, to see whether significant responses could be obtained. The child's reactions to the verbal material were evaluated and questions were asked to gain more material. The child was asked if he wanted to make up a story. Illustrative cases were given. The test was found to elicit valuable diagnostic data from the children. Revising the *Despert Fables*, Fine (35) found them of great use in providing information about the children's interpersonal relationships.

Castelnuovo-Tedesco (18) made a study of the relationship between handwriting and intelligence, originality, anxiety, compulsiveness, physical sex, and masculinity. Six judges rated 100 handwriting samples on a five-point scale for each of the variables. Coefficients of contingency between the ratings and outside criteria were significant at the 1 or 2 percent level of significance for all factors.

Further attempts were made to objectify projective technics. Fassett (34) attempted to revise Sargent's *Test of Insight into Human Motives* and to simplify the scoring. Sheriffs (75) devised a thirty-item "intuition questionnaire" with four variables for scoring. Autobiographies of the ninety-three college students who took the test were used as part of the validating criteria. Results were felt to be relatively quantitative, valid, and reliable.

The *Lowenfeld Mosaic Test* was used by McCullough and Girdner (52) with mental defectives and by Colm (22) as part of a battery of tests used in a child guidance clinic. Results were felt to be of diagnostic value.

Conclusion

Projective technics continued to develop in the past three years. Increasing efforts were made to improve on the reliability of the methods and to investigate their validity, their underlying hypotheses, and the effect of varying the procedure upon the fantasy product. Much more work along these lines would seem to be needed. More norms, both developmental and in terms of frequency, would be of value in evaluating fantasy material. More investigation of the relationship of projective technics to each other and to life history data would seem to be indicated.

Bibliography

1. ALBERMAN, NANETTE, and SCHAEFFER, VIRGINIA. "An Experiment in Story-telling." *Studies of Children*. (Meyer, Gladys E., editor.) New York: King's Crown Press, 1948. p. 26-58.
2. ALSCHULER, ROSE H., and HATTWICK, LABERTA W. *Painting and Personality: a Study of Young Children*. Two volumes. Chicago: University of Chicago Press, 1947. 590 p.
3. ALTUS, WILLIAM D. "Some Correlates of the Group Rorschach and the Schizophrenia Scale of the Group MMPI Among Two Groups of 'Normal' College Students." *Journal of Consulting Psychology* 12: 375-78; November-December 1948.
4. ANDERSON, IRMGARD, and MUNROE, RUTH L. "Personality Factors Involved in Student Concentration on Creative Painting and Commercial Art." *Rorschach Research Exchange and Journal of Projective Techniques* 12: 141-54; No. 3, 1948.
5. BACH, GEORGE R. "Father-Fantasies and Father-Typing in Father-Separated Children." *Child Development* 17: 63-80; March-June 1946.
6. BACH, GEORGE R., and BREMER, GLORIA. "Projective Father Fantasies of Pre-adolescent, Delinquent Children." *Journal of Psychology* 24: 3-17; July 1947.
7. BELL, JOHN E. "A Comparison of Children's Fantasies in Two Educated Projective Techniques." *American Psychologist* 3: 263; July 1948.
8. BELL, JOHN E. *Projective Technics: A Dynamic Approach to the Study of Personality*. New York: Longmans, Green and Company, 1948. 533 p.
9. BELLAK, LEOPOLD; EKSTEIN, RUDOLF; and BRAVERMAN, SYDELL. "A Preliminary Study of Norms for the Thematic Apperception Test." *American Psychologist* 2: 271; August 1947.
10. BETTELHEIM, BRUNO. "Self-Interpretation of Fantasy; the Thematic Apperception Test as an Educational and Therapeutic Device." *American Journal of Orthopsychiatry* 17: 80-100; January 1947.
11. BLAIR, GLEN M., and CLARK, RONALD W. "Personality Maladjustments of Ninth-Grade Pupils as Measured by the Multiple Choice Rorschach Test and the California Test of Personality." *Journal of Educational Psychology* 37: 13-20; January 1946.
12. BLUM, GERALD S. "A Study of the Psychoanalytic Theory of Psychosexual Development." *Genetic Psychology Monographs* 39: 3-99; February 1949.
13. BLUM, LUCILLE H., and DRAGOSITZ, ANNA. "Finger-Painting: the Developmental Aspects." *Child Development* 18: 88-105; September 1947.
14. BRONFENBRENNER, URIE, and NEWCOMB, THEODORE M. "Improvisations—an Application of Psychodrama in Personality Diagnosis." *Sociatry* 4: 367-82; March 1948.
15. BUROS, OSCAR K., editor. *The Third Mental Measurements Yearbook*. New Brunswick, New Jersey: Rutgers University Press, 1949. 1047 p.
16. CARP, FRANCES M. "Constriction as Rated on Three Productions." *American Psychologist* 3: 284; July 1948.
17. CARR, ARTHUR. "An Evaluation of Nine Nondirective Psychotherapy Cases by Means of the Rorschach." *Journal of Consulting Psychology* 13: 196-205; June 1948.
18. CASTELNUOVO-TEDESCO, PETER. "A Study of the Relationship Between Handwriting and Personality Variables." *Genetic Psychology Monographs* 37: 167-220; May 1948.
19. CHEIN, ISIDOR, and EVANS, MARY C. "The Movie Story Game: A Projective Test of Interracial Attitudes for Use with Negro and White Children." *American Psychologist* 3: 268-69; July 1948.
20. CLARK, JERRY H. "Some MMPI Correlates of Color Responses in the Group Rorschach." *Journal of Consulting Psychology* 12: 384-86; November-December 1948.
21. COLEMAN, WILLIAM. "The Thematic Apperception Test. I. Effect of Recent Experience. II. Some Quantitative Observations." *Journal of Clinical Psychology* 3: 257-64; July 1947.
22. COLM, HANNA. "The Value of Projective Methods in the Psychological Examination of Children: the Mosaic Test in Conjunction with the Rorschach and Binet Tests." *Rorschach Research Exchange and Journal of Projective Techniques* 12: 216-37; No. 4, 1948.

23. COMBS, ARTHUR W. "A Comparative Study of Motivations as Revealed in Thematic Apperception Studies and Autobiography." *Journal of Clinical Psychology* 3: 65-75; January 1947.
24. COMBS, ARTHUR W. "The Use of Personal Experience in Thematic Apperception Test Story Plots." *Journal of Clinical Psychology* 2: 357-63; October 1946.
25. COSTIN, FRANK, and EISERER, PAUL E. "Students' Attitudes Toward School Life as Revealed by a Sentence Completion Test." *American Psychologist* 4: 289; July 1949.
26. DORKEY, MARGARET, and AMEN, ELISABETH W. "A Continuation Study of Anxiety Reactions in Young Children by Means of a Projective Technique." *Genetic Psychology Monographs* 35: 139-83; May 1947.
27. DUBIN, ELISABETH R. "The Effect of Training on the Tempo of Development of Graphic Representation in Preschool Children." *Journal of Experimental Education* 15: 166-73; December 1946.
28. DUNKEL, HAROLD B. "The Effect of Personality on Language Achievement." *Journal of Educational Psychology* 38: 177-82; March 1947.
29. EISERER, PAUL E. "The Relative Effectiveness of Motion and Still Pictures as Stimuli for Eliciting Fantasy Stories About Adolescent-Parent Relationships." *Genetic Psychology Monographs* 39: 206-75; May 1949.
30. ELKISCH, PAULA. "The 'Scribbling Game'—a Projective Method." *Nervous Child* 7: 247-56; July 1948.
31. ENGLAND, ARTHUR O. "Non-structured Approach to the Study of Childhood Fears." *Journal of Clinical Psychology* 2: 364-68; October 1946.
32. ENGLE, THELBURN L. "The Use of the Harrower-Erickson Multiple Choice (Rorschach) Test in Differentiating Between Well-Adjusted and Maladjusted High-School Pupils." *Journal of Educational Psychology* 37: 550-56; December 1946.
33. ERON, LEONARD D. "Frequencies of Themes and Identifications in the Stories of Schizophrenic Patients and Non-Hospitalized College Students." *Journal of Consulting Psychology* 12: 387-95; November-December 1948.
34. FASSETT, KATHERINE K. "A Preliminary Investigation of the Sargent Test." *Journal of Clinical Psychology* 4: 45-46; January 1948.
35. FINE, REUBEN. "Use of the Despert Fables (Revised Form) in Diagnostic Work with Children." *Rorschach Research Exchange and Journal of Projective Techniques* 12: 106-18; No. 2, 1948.
36. FORD, MARY. *The Application of the Rorschach Test for Young Children*. Child Welfare Monograph Series, No. 23. Minneapolis: University of Minnesota Press, 1946. 114 p.
37. FRANK, LAWRENCE K. *Projective Methods*. Springfield, Illinois: C. C. Thomas, 1948. 86 p.
38. GOODENOUGH, FLORENCE L. "The Appraisal of Child Personality." *Psychological Review* 56: 123-31; May 1949.
39. GOTHBERG, LAURA C. "A Comparison of the Personality of Runaway Girls with a Control Group as Expressed in the Themas of Murray's Thematic Apperception Test." *American Journal of Mental Deficiency* 51: 627-31; April 1947.
40. GROVER, KATHLEEN B. "Use of English Compositions To Gain Understanding of Pupils." *School Review* 54: 605-10; December 1946.
41. GUSTAV, ALICE. "Estimation of Rorschach Scoring Categories by Means of an Objective Inventory." *Journal of Psychology* 22: 253-60; October 1946.
42. HAMLIN, ROY M., and ALBEE, GEORGE W. "Muench's Tests Before and After Non-directive Therapy: a Control Group for His Subjects." *Journal of Consulting Psychology* 12: 412-16; November-December 1948.
43. HARMS, ERNEST. "Play Diagnosis: Preliminary Considerations for a Sound Approach." *Nervous Child* 7: 233-46; July 1948.
44. HAVIGHURST, ROBERT J., and TABA, HILDA. *Adolescent Character and Personality*. New York: John Wiley and Sons, 1949. 315 p.
45. JOEL, WALTHER. "The Use of the Make a Picture Story (MAPS) Test with Disturbed Adolescents." *Rorschach Research Exchange and Journal of Projective Techniques* 12: 155-64; No. 3, 1948.
46. JOLLES, ISAAC. "The Diagnostic Implications of Rorschach's Test in Case Studies of Mental Defectives." *Genetic Psychology Monographs* 36: 89-197; November 1947.

47. KABACK, GOLDIE R. *Vocational Personalities; an Application of the Rorschach Group Method*. Contributions to Education, No. 924. New York: Teachers College, Columbia University, 1946. 116 p.
48. KELLY, E. LOWELL. "Research in the Selection of Clinical Psychologists." *Journal of Clinical Psychology* 3: 39-42; January 1947.
49. LAWSHE, CHARLES H., JR., and FORSTER, MAX H. "Studies in Projective Techniques: I. The Reliability of a Multiple Choice Group Rorschach Test." *Journal of Applied Psychology* 31: 199-211; April 1947.
50. McCANDLESS, BOYD R. "The Rorschach as a Predictor of Academic Success." *Journal of Applied Psychology* 33: 43-50; February 1949.
51. McCLELLAND, DAVID C., and OTHERS. "The Projective Expression of Needs. IV. The Effect of the Need for Achievement on Thematic Apperception." *Journal of Experimental Psychology* 39: 242-55; April 1949.
52. McCULLOUGH, THOMAS L., and GIRDNER, JOHN B. "Use of the Lowenfeld Mosaic Tests with Mental Defectives." *American Journal of Mental Deficiency* 53: 486-96; January 1949.
53. MEISTER, DAVID. "Adjustment of Children as Reflected in Play Performance." *Journal of Genetic Psychology* 73: 141-55; September 1948.
54. MERRILL, MAUD A. *Problems of Delinquency*. Boston: Houghton Mifflin Company, 1947. 403 p.
55. MONTALTO, FANNIE D. "An Application of the Group Rorschach Technic to the Problem of Achievement in College." *Journal of Clinical Psychology* 2: 254-60; July 1946.
56. MUENCH, GEORGE A. *An Evaluation of Nondirective Psychotherapy*. Applied Psychology Monographs, No. 13. Stanford University, California: Stanford University Press, 1947. 168 p.
57. MUNROE, RUTH L. "Rorschach Findings on College Students Showing Different Constellations of Subscores on the A.C.E." *Journal of Consulting Psychology* 10: 301-16; November-December 1946.
58. MUNROE, RUTH L. "The Use of Projective Methods in Group Testing." *Journal of Consulting Psychology* 12: 8-15; January-February 1948.
59. NAPOLI, PETER J. "A Finger Painting Record Form." *Journal of Psychology* 26: 31-43; July 1948.
60. NAUMBERG, MARGARET. *Studies of the "Free" Art Expression of Behavior Problem Children and Adolescents as a Means of Diagnosis and Therapy*. Nervous and Mental Disease Monographs, No. 71. New York: Cooledge Foundation, 1947. 225 p.
61. NORTHWAY, MARY L., and WIGDOR, BLOSSOM T. "Rorschach Patterns Related to the Sociometric Status of School Children." *Sociometry* 10: 186-99; May 1947.
62. PHILLIPS, EDWIN, and STROMBERG, ELEROY L. "A Comparative Study of Finger-painting Performance in Detention Home and High-School Pupils." *Journal of Psychology* 26: 507-15; October 1948.
63. RADKE, MARIAN J. *The Relation of Parental Authority to Children's Behavior and Attitudes*. University of Minnesota Institute of Child Welfare Monographs, No. 22. Minneapolis: University of Minnesota Press, 1946. 123 p.
64. ROBINSON, ELIZABETH F. "Doll Play as a Function of the Doll Family Constellation." *Child Development* 17: 99-119; September 1946.
65. ROSENZWEIG, SAUL; FLEMING, EDITH E.; and CLARKE, HELEN J. "Revised Scoring Manual for the Rosenzweig Picture-Frustration Study." *Journal of Psychology* 24: 165-208; October 1947.
66. ROSENZWEIG, SAUL; FLEMING, EDITH E.; and ROSENZWEIG, LOUISE. "The Children's Form of the Rosenzweig Picture-Frustration Study." *Journal of Psychology* 26: 141-91; July 1948.
67. ROSENZWEIG, SAUL, and KOGAN, KATE LEVINE. *Psychodiagnosis: An Introduction to Tests in the Clinical Practice of Psychodynamics*. New York: Grune and Stratton, 1949. 380 p.
68. ROTTER, JULIAN B. "The Present Status of the Rorschach in Clinical and Experimental Procedures." *Journal of Personality* 16: 304-11; March 1948.
69. ROTTER, JULIAN B., and WICKENS, DELOS D. "The Consistency and Generality of 'Social Aggressiveness' Made from Observations of Role Playing Situations." *Journal of Consulting Psychology* 12: 234-39; July-August 1948.

70. SARASON, SEYMOUR B., and SARASON, ESTHER K. "The Discriminatory Value of a Test Pattern with Cerebral Palsied, Defective Children." *Journal of Clinical Psychology* 3: 141-47; April 1947.
71. SCHAFER, SARAH, and LEITCH, MARY. "An Exploratory Study of the Usefulness of a Battery of Psychological Tests with Nursery School Children." *American Journal of Psychiatry* 104: 647-52; April 1948.
72. SCHMIDL, FRITZ. "The Rorschach Test in Juvenile Delinquency Research." *American Journal of Orthopsychiatry* 17: 151-60; January 1947.
73. SEARS, ROBERT R. "Influence of Methodological Factors in Doll Play Performance." *Child Development* 18: 190-97; December 1947.
74. SEARS, ROBERT R.; PINTLER, MARGARET H.; and SEARS, PAULINE R. "Effect of Father Separation on Preschool Children's Doll Play Aggression." *Child Development* 17: 219-43; December 1946.
75. SHERIFFS, ALEX C. "The 'Intuition Questionnaire': a New Projective Technique." *Journal of Abnormal and Social Psychology* 43: 326-37; July 1948.
76. SIEGEL, MIRIAM G. "The Diagnostic and Prognostic Validity of the Rorschach Test in a Child Guidance Clinic." *American Journal of Orthopsychiatry* 18: 119-33; January 1948.
77. STEIN, MORRIS I. *The Thematic Apperception Test*. Cambridge, Massachusetts: Addison-Wesley Press, 1948. 95 p.
78. STEINER, MATILDA E. "The Use of Projective Techniques in Industry (A Review of Published Material)." *Rorschach Research Exchange and Journal of Projective Techniques* 12: 171-74; No. 3, 1948.
79. SYMONDS, PERCIVAL M. *Adolescent Fantasy*. New York: Columbia University Press, 1949. 397 p.
80. SYMONDS, PERCIVAL M. *Symonds Picture-Story Test*. New York: Bureau of Publications, Teachers College, Columbia University, 1948. 20 plates; manual, 20 p.
81. THOMPSON, GRACE M. "College Grades and the Group Rorschach." *Journal of Applied Psychology* 32: 398-407; August 1948.
82. THOMPSON, GRACE M. "MMPI Correlates of Certain Movement Responses in the Group Rorschachs of Two College Samples." *Journal of Consulting Psychology* 12: 379-83; November-December 1948.
83. TOMKINS, SILVAN S. *The Thematic Apperception Test; the Theory and Technique of Interpretation*. New York: Grune and Stratton, 1947. 297 p.
84. OFFICE OF STRATEGIC SERVICES, ASSESSMENT STAFF. *Assessment of Men*. New York: Rinehart and Company, 1948. 541 p.
85. WITTENBORN, JOHN R. "Some Thematic Apperception Test Norms and a Note on the Use of the Test Cards in the Guidance of College Students." *Journal of Clinical Psychology* 5: 157-61; April 1949.
86. WOLFF, WERNER. *The Personality of the Preschool Child*. New York: Grune and Stratton, 1946. 341 p.
87. WYATT, FREDERICK. "The Scoring and Analysis of the Thematic Apperception Test." *Journal of Psychology* 24: 319-30; October 1947.

CHAPTER VI

Measurement of Educational Achievement in the Schools

WARREN G. FINDLEY AND ALLAN B. SMITH

SINCE 1938, when educational tests were last reviewed in the same issue with psychological tests, a number of milestones in educational measurement have been passed. In the meantime, readers of the REVIEW have been apprised of these developments in each cycle by chapters on Tests and Measurements in the issues on Methods of Research and Appraisal in Education. The readers of this chapter are specifically directed to the chapter by Schrader and Conrad in the December 1948 REVIEW for independent analysis and comment on some of the items reviewed here. Because validation studies and applications of achievement tests often include validation and application of tests of intelligence, aptitude, and personality, readers are also advised to consult the several other chapters of this issue to which discussion of certain articles has occasionally had to be arbitrarily assigned.

Critical Articles on Measurement Procedure

In discussing the problems involved in appraising a school, Harris (33) pointed to the ultimate necessity of measuring effects on students through appropriate tests rather than by analyzing school facilities; he further pointed out that measurement of effects implies measurement of gains rather than evaluation of present status. Brownell (11) offered a criticism of the criteria of learning implicit in most educational measurement. He insisted that we raise our sights from measures of rate and accuracy of performance to measures of level of process used, from evidence of immediate gains to that of more permanent gains, and from ability to use learning in closely similar situations to transferability to essentially new situations, especially after a significant lapse of time.

Simpson (68) formulated a series of sixteen questions to guide interpretation of test results by the school people who are the basic users of achievement tests. These questions ask proper consideration of policies of retardation and elimination, motivation of teachers and students in preparing for tests, length of school term, repetition of tests, and cultural advantages and disadvantages of the students, among others. Rinsland (62) described and included a form for evaluating standardized tests for students of measurement, test purchasers, etc. Swenson stated that "A great deal of the present skepticism concerning standardized tests might be removed or at least tempered if those who use them would seriously reconsider the question:

When is test performance good enough?" (71, p. 115). She feels that users of standardized tests commonly ignore performance of the individual child, neglect to study pupil performance on specific test items, and generally fail to make full use of the test results. In an article on use of tests in counseling, Traxler (75) pointed out that altho research data on the prognostic value of tests designed for individual appraisal are inadequate and there is need for a common scale into which the results of all tests can be translated, the main need at present is to extend and improve the use which counselors are making of the available procedures.

Wrightstone (88) observed that during the past three decades measurement of achievement has progressed from simple and crude to relatively refined measures. Two major trends were pointed out: (a) a growing concern about the appraisal of such objectives as critical thinking and work-study skills to supplement such established objectives as acquisition of information, concepts, and basic skills; and (b) an increasing use of factor analysis of the component abilities of achievement within various subject-matter areas. Current problems include the construction of exercises to measure some of the so-called intangible objectives, factor analysis of the components of various aptitudes and achievements, improved methods of estimating validity and reliability of tests, establishment of better norms, and improved methods of selecting samples for standardization purposes. Scates (66), in his review of educational measurement and research, remarked that he has been more critical of the use of educational tests than of the instruments themselves, for in practice, tests are often taken to represent the whole of the educational goal. He further stated that the modern measurement movement was not developed by teachers or for teaching purposes and consequently does not supply teachers with goals or effective aids, but rather was developed for and by research workers to answer basic generalized questions.

Dexter (21) emphasized the following points: (a) examinations and methods of marking determine what and how students learn; (b) there is need for a sociology of testing and marking; (c) there should be interaction between examining methods and objectives of the course; and (d) examinations should be on the subject or skill, not on the course. Freeman (33) stated his conviction that the use of objective examinations is greatly overdone, which is bound to have a harmful effect on study and learning in America; that objective tests have their uses; that the essay test should be used to a greater extent; and that the free expression of thought thru language should be restored to the position of dignity which it deserves.

Engelhart (25) offered a number of very practical guiding principles to be recommended to teachers collaborating in the production of tests for use in large-scale programs that involve the use of machine-scorable answer sheets. Hammock, in a report surveying teacher-made achievement tests, listed three categories of poor items: (a) uncomprehended labels, (b) definitions detached from their use, and (c) items of questionable

value. He concluded, "There was no premeditated malice in the poor examinations—there just was not enough meditation" (37, p. 396).

Surveys and Programs

Otto (59) found that 73 percent of 286 non-Texas public schools and 74 percent of 46 campus demonstration schools used standardized achievement tests once or twice a year and that 20 percent of the public schools and 45 percent of the demonstration schools gave achievement tests semi-annually. In a survey of testing practices in California, Belanger (8) gave the percentage of school systems using each of several instruments. In more than 25 percent, scoring was done by clerks or machine; in 66 percent, the most vexing problem indicated was effective use of test results.

Durost (23) and Findley (28) described their recommendations for group testing programs for school systems, and Seymour (67) described the program currently in operation in Rochester, New York.

Elementary-School Studies and Reports

General. In a special issue devoted to testing in the elementary school, Hildreth (43) contrasted the standardized test and the teacher-made classroom achievement test with reference to the appropriate use that should be made of each and their corresponding advantages and disadvantages. Findley (27) proposed principles for securing economic efficiency and a sound psychological impact from an elementary-school testing program.

Reading. Triggs (78) described the work of the Committee on Diagnostic Reading Tests in preparing the *Survey Test* and *Diagnostic Tests* of the recently published comprehensive battery. In response to citizens' criticisms that basic skills are today being taught poorly, Finch and Gillenwater (26) contrasted reading proficiency of two groups of sixth-grade pupils in a citywide system, one recently in school and the other in school seventeen years earlier. They concluded that the present teaching of reading was more successful in producing the outcomes measured than was that of the earlier period. Carlson (12) analyzed test data from fifth-grade pupils and concluded that accuracy of comprehension of slow and fast readers was dependent on several factors and that no consistent or significantly large relationships existed between intelligence and reading efficiency, speed, or accuracy. Chall (13) inferred from testing sixth- and eighth-grade children that greater knowledge of the subject area in which the reading occurred resulted in increased comprehension of reading. In a review of thirty-seven articles, Hildreth (42) found evidence of a significant relationship between reading achievement and other language arts.

Townsend (74) reported correlations of .46 to .70 between scores in spelling and in reading comprehension and vocabulary and only slightly

lower correlations between scores in spelling and academic aptitude for pupils in Grades III thru XII. Russell (64) found correlations of .66 to .86 among such abilities as spelling, word recognition, reading comprehension, word meaning, and reading speed and concluded that improvement in one of these abilities should be reflected in improvement in the others. Kyte (48) found that some elementary-school pupils could be excused from further instruction in spelling, but stated that they should have periodic testing to determine if they were continuing development in spelling.

Malter (51, 52, 53) studied the ability of 376 children in Grades IV to VIII in reading cross-sections, process-diagrams, and conventionalized diagrammatic symbols. Findings: wide variability, superiority of bright children, and great improvability in reading cross-sections; reading of process-diagrams greatly improved by labeling; an automatic tendency to read from upper left; inconclusive trends in ability to interpret symbols. Gates (35) found higher intercorrelations between tests involving similar diagnostic procedures (e.g., flash exposure) but different material (e.g., words vs. phrases) than between tests based on the same stimulus materials (e.g., words) but different procedures (e.g., flash vs. blending) to an extent that appeared to justify the claim that the procedures diagnosed distinguishable skills of reading.

Miller (55) reported that a background of radio programs during the testing period did not materially reduce reading-test scores of sixth- and seventh-grade children. Mitchell (56) reported similarly that the background of a radio variety program did not reduce reading-test scores of sixth-grade pupils with I.Q.'s over 100, that it did reduce the scores of those with I.Q.'s below 100, and that a background of musical programs actually increased significantly the scores of pupils with I.Q.'s above 100. Sperzel (69) concluded that the reading of comic books had no appreciable effect upon growth in vocabulary and reading comprehension of fifth-grade pupils. Furthermore, Heisler (39) found no significant differences in school achievement between elementary-school children who did and those who did not attend moving pictures, read comic books, and listen to serial radio programs to an excess.

Geography. In a study based on 1055 fifth-grade pupils, Tiedeman (72) found review tests in geography helpful in the manner of other memory devices if used frequently immediately after learning and less frequently with the passage of time.

Arithmetic. Olander, Van Wagenen, and Bishop (58) reported construction of an instrument to measure arithmetic ability of first-grade children and its use to predict arithmetic achievement three years later. Searching for reasons for the lower achievement of mentally retarded children, Cruickshank (16) discovered that the inferiority in achievement is intensified if the arithmetic problem contains extraneous data. Ramharter and Johnson (61) studied the work of good and poor pupils in sixth-grade arithmetic to determine the characteristics of their procedures. Studying the retention and relearning of arithmetic skills in Grades VII and VIII,

Davis and Rood (20) found that there were steady gains in average scores and that problems once solved correctly and missed on the second testing were largely solved correctly on the third testing. This phenomenon, they concluded, reflects the importance of relearning that accompanies further study of advanced topics dependent on skills learned earlier. In an experiment designed to change the profiles of pupils taking achievement tests, Tilton (73) instructed teachers to concentrate in order to peak the arithmetic ordinate of the experimental group of pairs of students matched for chronological, mental, and arithmetic ages. The arithmetic ordinates of the profiles were raised a significant amount for some pupils.

Secondary-School Studies and Reports

General. The National Association of Secondary-School Principals devoted the December 1948 issue of its *Bulletin* to King's interpretive guide for using tests (46). Traxler (76) surveyed the status of paper-and-pencil reading tests and analyzed some of the current tests.

Bender and Davis (9) obtained students' opinions of testing practices in Colorado secondary schools and found that students preferred essay tests, of some difficulty, stressing application rather than knowledge, at weekly intervals, with advance notice, scored by teachers, with comments or corrections indicated. Vallance (81), studying seniors, presented inconclusive results of the differential effects of studying for and taking essay tests as distinguished from objective tests.

Lobaugh (49), duplicating previous reports, tested high-school seniors with the *Myers-Ruch High School Progress Test* and found that altho the boys were receiving lower course marks, the median of their scores on the test was ten points higher than the median of the scores of the girls.

Reading. Artley (4) reported a coefficient of correlation of .75 between reading comprehension in general and in specific areas, computed from scores on tests in each administered to eleventh-grade pupils. The relationship between achievement and reading is not a simple one, according to Aukerman (6), who found, by testing several general and specific abilities and skills, that general reading ability was the most significant differentiating factor between good and poor eleventh-grade pupils. Eagle (24) concluded from his research with ninth-grade pupils that improved mathematics proficiency did not necessarily follow improved general reading proficiency but that more attention should be paid to the reading of specific concepts in mathematics.

Mathematics. In a discussion of evaluating achievement in mathematics, Traxler (77) concluded that there should be coordination of evaluation procedures, that no one testing program fits all schools, that tests should be selected in the light of the objectives and curriculum organization of the school, and that a principal educational need at present is to make objective tests less static and more responsive to desirable innovations in the curriculum. Sister Mary Miriam Ryan (65) administered intelli-

gence and mathematics achievement tests to ninth-grade pupils and reported the correlations between capacity and accomplishment in general mathematics were higher for the upper levels of capacity and lower for the lower levels of capacity.

Miscellaneous. Daly, Brugger, and Anderson (19) appraised the gains made in a fusion course of English and social studies (largely American history) implied in the educational film, *Land of Liberty*. Significant gains were reported on most parts of the tests in English, reading, and American history. Baten and Hatcher (7) reported that eleventh-grade girls in home economics classes did approximately as well as did twelfth-grade girls in similar classes on a 250-item test following a four-week unit on consumer buying. Gauger (36) found that high-school students evaluated their fellow students on performance in a speech course more leniently than did teachers, but that they did not disagree materially with the teacher in the order of evaluation.

University Studies and Reports

General. Watson (83) surveyed final examination practices of twenty-nine universities and sixty-four colleges, finding that 75 percent or more of these institutions used between two and six days for final examinations, set two to four examination periods per day, and allowed two or three hours for each examination. Troyer (80) described eleven types of advisory service provided the faculty of Syracuse University by the Evaluation Service Center. Thruout, a nonauthoritarian advisory and service attitude is maintained. Diederich (22) reported on the comprehensive examination system at the University of Chicago. Wieman (87) described the Antioch terminal integrating examination.

Super, Braasch, and Shay (70) reported a study in which a variety of distractions normally avoided by examiners were applied to a group taking clerical and intelligence tests. Only conflicting, unreliable differences from a control group were found. Henderson, Crews, and Barlow (40) reported that classical music did not distract but that popular music did distract students significantly on the paragraph section of a reading test.

Assum and Levy (5) compared a group of normal students with a group known to be nonadjusted and found no significant difference in scholastic aptitude, tho the mean achievement of the adjusted group was higher. McCurdy (50) found that the basal metabolic rates of thirty women students bore correlations of .43 and .53 to grades in a psychology course and a three-semester point-hour ratio, respectively. Because of the low correlation (.06) of basal metabolic rate with intelligence, addition of the intelligence factor produced multiple correlations with grades of .69 and .71, respectively.

Krueger (47) purposely introduced grading errors into weekly quizzes and found that more than 90 percent of the students reported the error when its effect was to lower their grades but made no report when the

effect of the error was to raise their grades, altho they were urged to report all errors in *both* directions. When the students were shown that the instructor had an accurate record of grades and grading errors, the percentages of students reporting corrections unfavorable to themselves rose from 10 to 99. Weinland (85) reported administering two forms of tests, both of which contained in scrambled fashion items in common with the other form and also items not in common with the other form. His conclusion from the results obtained in the experiment was that such cribbing as did take place had little effect on grades.

Comrey (15) applied factor analysis to grades received in eight subjects at West Point and scores earned on twelve classification tests and one apparatus test administered to 815 cadets in the class of 1946. One of the eight factors found, "academic achievement," may have involved a halo of subjective judgment of teachers, but most certainly it involved whatever it takes to succeed in course work as opposed to tests, since all academic grades had substantial loadings on the factor while no test had such loading.

Welborn (86) found that veterans excelled regular students at a teachers college by small amounts in almost all courses for which comparable data were available, especially in professional courses. Gains from prewar to postwar by veterans whose schooling had been interrupted were substantial, especially for those with low prewar grades. Garnezy and Crose (34) matched veterans and nonveterans at the University of Iowa on the basis of their percentile rank on the *Tests of General Educational Development* and reported that the veterans averaged slightly higher (.10 grade point) in academic achievement than did comparable non-veterans.

Charles (14) found that students in the top quarter of their high-school classes on entering the University of Nebraska were three years younger than average, excelled on achievement tests, made consistently slightly higher college grades, were represented in much greater proportions in fourth-year classes than average or low-quarter groups, and won a highly disproportionate share of academic honors.

Angell (3) reported the effect on students in college freshman chemistry of knowing immediately the results of their quizzes. He used for a control group regular machine-scored answer sheets which were returned at the next class meeting and for the experimental group the *Angell-Troyer Punchboard*. On the final examination in the course the difference between the equated groups was significant at the 1 percent level in favor of the experimental group. Jones and Sawyer (45) reported use of the *Angell-Troyer Punchboard* in a freshman course at Syracuse University. Results of the experimental group tested by the punchboard compared with those of the matched group tested by conventional method closely approached significance ($t = 1.68$).

Reading. Anderson and Morse (2) found that reading scores of 590 veteran students at the University of Michigan increased as the testing

progressed from freshman to senior students and, more remarkably, that scores increased with the number of years away from school, independent of class in college. The latter fact suggests that veterans farther removed from schooling may have tended to set themselves a higher standard of basic proficiency before undertaking the college study under GI benefits. The authors suggest general maturity as the chief factor. Ammons and Hieronymus (1) found it possible to produce statistically and practically significant gains of an experimental group over a control group on reading speed and reading comprehension as a result of an intensive twenty-hour reading improvement program. Robinson (63) found that students trained in reading insurance papers excelled untrained students even when the difficulty of the vocabulary was eliminated as a factor in reading. He concluded that language structure, as well as vocabulary, is a factor in reading difficulty.

Flesch (31) described and illustrated a simplified readability approach that yields separate measures of reading ease and human interest, each on a scale from 0 to 100. These are for use in courses in composition, creative writing, journalism, and advertising as well as in the applied fields themselves. The new element is percentage of "personal sentences," defined as spoken sentences with or without quotations, questions, commands, requests, exclamation, etc. The two separate measures were shown to give relatively independent estimates of two significant aspects of readability, the first showing a correlation (.70) almost as high as the original single formula did (.74), while the second showed a correlation of .43. Flesch (29, 30) incorporated his earlier and later readability formulas in full-length books published in 1946 and 1949, respectively. Dale and Chall (17, 18) reported a readability formula based on two factors: (a) average length of sentence and (b) proportion of words outside a 3000-word list.

Watts (84) described a group-study approach to the improvement of reading in which the students studied clinical cases which they found helpful as a basis for analyzing their own disabilities. Both clinical and statistical evidence indicated substantial improvement in most cases and at least considerable insight into their difficulties in all cases.

Language Arts. Votaw (82) reported that a *Library and Study Materials Test* was slightly the best of three predictors of freshman grades in a class of 412 at Southwest Texas State College. He also found that students admitted to college in 1945 were inferior in library skills to those admitted in 1942 altho they were about equal on the other two tests. This situation he attributed to the great inroads made on librarians and library service in high schools during the war.

Science, Engineering, Mathematics. Hendricks (41) analyzed several thousand test items used in courses in college chemistry. He found a great disparity between the relative emphasis professed to be given to certain outcomes of instruction and the percentage of test items related to them. Mason (54) reported on the areas in which college students showed greatest proficiency at the beginning and at the end of a biology course.

Pierson (60), using 463 prewar graduates, reported that four-year college averages in engineering correlated .58 with high-school average, .67 with first-quarter college average, .75 with first-year college average, and .43 with faculty estimate of future success in professional practice. Utilizing these data he upheld the validity of marks in general. The possibility of contamination of estimates of professional success by knowledge of academic achievement in college was not mentioned. Frederiksen (32) found a multiple correlation of .63 between college freshman mathematics grades and a combination of general standing in high-school class and score on the mathematical portion of the College Entrance Examination Board's *Scholastic Aptitude Test*, virtually identical regression lines being found for 170 veterans and 250 nonveterans. The *Cooperative Survey Test in Mathematics* was equally good as a single predictor but had more in common with the two measures than they had with each other, so it added little to the multiple correlation.

Miscellaneous. Horrocks and Troyer (44) described the detailed planning, construction, and validation of three one-hour case-study tests of ability to use knowledge of human growth and development. Each test had a reliability of approximately .75 with part-scores for diagnosis and remediation almost as reliable as the totals. Intercorrelations of the tests were .55, .39, and .62, which may be considered to be better estimates of their validity than is a correlation of .38 for one of them and grades in a course in adolescent development. Nahm (57) found statistically reliable differences among students of twelve Minnesota schools of nursing with regard to knowledge of mental hygiene. She tended to attribute these differences to differences in mental hygiene atmospheres in the general administration of prenursing as well as nursing institutions. Troyer (79) described and illustrated an objective comprehensive examination developed to measure the desired outcomes of a program leading to the M.S. in education. The examination sought to measure, for four subject areas: knowledge of facts and principles, ability to interpret professional data, ability to make decisions in professional situations, and knowledge of current professional literature. Bottorf (10) compared the gains in art appreciation of groups taught by lecture and by creative activity with materials supplemented by lectures and found that the difference in favor of the latter more inclusive method was not statistically reliable.

Bibliography

1. AMMONS, ROBERT B., and HIERONYMUS, ALBERT N. "Critical Evaluation of a College Program for Reading Improvement." *Journal of Educational Psychology* 38: 449-70; December 1947.
2. ANDERSON, IRVING H., and MORSE, WILLIAM C. "The Reading of Veterans." *Journal of Higher Education* 17: 375-77; October 1946.
3. ANGELL, GEORGE W. "Effect of Immediate Knowledge of Quiz Results on Final Examination Scores in Freshman Chemistry." *Journal of Educational Research* 42: 391-94; January 1949.
4. ARTLEY, A. STERL. "General and Specific Factors in Reading Comprehension." *Journal of Experimental Education* 16: 181-86; March 1948.
5. ASSUM, ARTHUR L., and LEVY, SIDNEY J. "Comparative Study of Academic Ability and Achievement of Two Groups of College Students." *Journal of Educational Psychology* 38: 307-10; May 1947.
6. AUKERMAN, ROBERT C. "Differences in the Reading Status of Good and Poor Eleventh Grade Students." *Journal of Educational Research* 41: 498-515; March 1948.
7. BATEN, WILLIAM D., and HATCHER, HAZEL M. "Testing for Grade and School Differences Among the Scores of Home Economics Students." *Journal of Experimental Education* 16: 176-80; March 1948.
8. BELANGER, LAURENCE L. "Testing in California Schools." *California Journal of Secondary Education* 22: 108-11; February 1947.
9. BENDER, WILLIAM, JR., and DAVIS, ROBERT A. "What High School Students Think About Teacher-Made Examinations." *Journal of Educational Research* 43: 58-65; September 1949.
10. BOTTORF, EDNA A. "A Study Comparing Two Methods of Developing Art Appreciation with College Students." *Journal of Educational Psychology* 38: 17-44; January 1947.
11. BROWNELL, WILLIAM A. "Criteria of Learning in Educational Research." *Journal of Educational Psychology* 39: 170-82; March 1948.
12. CARLSON, THORSTEN R. "The Relationship Between Speed and Accuracy of Comprehension." *Journal of Educational Research* 42: 500-12; March 1949.
13. CHALL, JEANNE S. "The Influence of Previous Knowledge on Reading Ability." *Educational Research Bulletin* 26: 225-46; December 10, 1947.
14. CHARLES, DON C. "College Performance of Top-Quarter High-School Graduates." *Journal of Educational Psychology* 39: 82-91; February 1948.
15. COMREY, ANDREW L. "A Factorial Study of Achievement in West Point Courses." *Educational and Psychological Measurement* 9: 193-209; Summer 1949.
16. CRUICKSHANK, WILLIAM F. "Arithmetic Ability of Mentally Retarded Children: I. Ability To Differentiate Extraneous Materials from Needed Arithmetical Facts." *Journal of Educational Research* 42: 161-70; November 1948.
17. DALE, EDGAR, and CHALL, JEANNE S. "A Formula for Predicting Readability." *Educational Research Bulletin* 27: 11-20; January 21, 1948.
18. DALE, EDGAR, and CHALL, JEANNE S. "A Formula for Predicting Readability: Instructions." *Educational Research Bulletin* 27: 37-54; February 17, 1948.
19. DALY, GRACE M.; BRUGGER, HARMONY E.; and ANDERSON, KENNETH E. "America: An Experiment." *Journal of Educational Research* 41: 222-29; November 1947.
20. DAVIS, ROBERT A., and ROOD, EDWARD J. "Remembering and Forgetting Arithmetical Abilities." *Journal of Educational Psychology* 38: 216-22; April 1947.
21. DEXTER, LEWIS A. "Examinations as Instruments of, and Obstacles to, General Education." *School Review* 55: 534-41; November 1947.
22. DIEDERICH, PAUL B. "The Effects of Independent Comprehensive Examinations." *School Review* 55: 526-33; November 1947.
23. DUROST, WALTER N. "What Constitutes a Minimal School Testing Program?" *Educational and Psychological Measurement* 7: 45-60; Spring 1947.
24. EAGLE, EDWIN. "The Relationship of Certain Reading Abilities to Success in Mathematics." *Mathematics Teacher* 41: 175-79; April 1948.
25. ENGELHART, MAX D. "Suggestions for Writing Achievement Exercises To Be Used in Tests Scored on the Electric Scoring Machine." *Educational and Psychological Measurement* 7: 357-74; Autumn 1947.

26. FINCH, FRANK H., and GILLENWATER, VIRGIL W. "Reading Achievement Then and Now." *Elementary School Journal* 49: 446-54; April 1949.
27. FINDLEY, WARREN G. "Getting Full Value from the Group Testing Program." *National Elementary Principal* 25: 20-23; June 1946.
28. FINDLEY, WARREN G. "A Group Testing Program for the Modern School." *Educational and Psychological Measurement* 5: 173-79; Summer 1945.
29. FLESCH, RUDOLF. *The Art of Plain Talk*. New York: Harper and Brothers, 1946. 210 p.
30. FLESCH, RUDOLF. *The Art of Readable Writing*. New York: Harper and Brothers, 1949. 237 p.
31. FLESCH, RUDOLF. "A New Readability Yardstick." *Journal of Applied Psychology* 32: 221-33; June 1948.
32. FREDERIKSEN, NORMAN. "Predicting Mathematics Grades of Veteran and Non-veteran Students." *Educational and Psychological Measurement* 9: 73-88; Spring 1949.
33. FREEMAN, FRANK N. "The Monopoly of Objective Tests." *Educational Forum* 10: 389-95; May 1946.
34. GARMETZ, NORMAN, and CROSE, JEAN M. "A Comparison of the Academic Achievement of Veteran and Non-Veteran Freshmen at the University of Iowa." *Journal of Educational Research* 41: 547-50; March 1948.
35. GATES, ARTHUR I. "A Correlational Study of a Battery of Reading Diagnostic Tests." *Journal of Educational Research* 40: 436-47; February 1947.
36. GAUGER, PAUL W. "A Comparison of the Ability of High-School Speech Students and Speech Experts in Rating a Speech Performance." *Journal of Educational Research* 42: 209-17; November 1948.
37. HAMMOCK, ROBERT C. "Think Before You Test." *Clearing House* 21: 396-99; March 1947.
38. HARRIS, CHESTER W. "The Appraisal of a School—Problems for Study." *Journal of Educational Research* 41: 172-82; November 1947.
39. HEISLER, FLORENCE. "A Comparison Between Those Elementary School Children Who Attend Moving Pictures, Read Comic Books and Listen to Serial Radio Programs to an Excess with Those Who Indulge in These Activities Seldom or Not at All." *Journal of Educational Research* 42: 182-90; November 1948.
40. HENDERSON, MACK T.; CREWS, ANNE; and BARLOW, JOAN. "A Study of the Effect of Music Distraction on Reading Efficiency." *Journal of Applied Psychology* 29: 313-17; August 1945.
41. HENDRICKS, BERNARD C. "How Good Are Our Tests?" *School Science and Mathematics* 47: 470-74; May 1947.
42. HILDRETH, GERTRUDE H. "Interrelations Among the Language Arts." *Elementary School Journal* 48: 538-49; June 1948.
43. HILDRETH, GERTRUDE H. "Standard Achievement Tests and Classroom Examinations." *National Elementary Principal* 25: 27-30; February 1946.
44. HORROCKS, JOHN E., and TROYER, MAURICE E. "Case Study Tests of Ability To Use Knowledge of Human Growth and Development." *Educational and Psychological Measurement* 7: 23-36; Spring 1947.
45. JONES, HOWARD L., and SAWYER, MICHAEL O. "A New Evaluation Instrument." *Journal of Educational Research* 42: 381-85; January 1949.
46. KING, JOSEPH E. "Using Tests in the Modern Secondary School." *Bulletin of the National Association of Secondary-School Principals* 32: 3-92; December 1948.
47. KRUEGER, WILLIAM C. F. "Students' Honesty in Correcting Grading Errors." *Journal of Applied Psychology* 31: 533-35; October 1947.
48. KYTE, GEORGE C. "When Spelling Has Been Mastered in the Elementary School." *Journal of Educational Research* 42: 47-53; September 1948.
49. LOBAUGH, DEAN. "Girls and Grades." *School Science and Mathematics* 47: 763-77; December 1947.
50. MCCURDY, HAROLD G. "Basal Metabolism and Academic Performance in a Sample of College Women." *Journal of Educational Psychology* 38: 363-72; October 1947.
51. MALTER, MORTON S. "The Ability of Children To Read a Process-Diagram." *Journal of Educational Psychology* 38: 290-98; May 1947.
52. MALTER, MORTON S. "The Ability of Children To Read Conventionalized Diagrammatic Symbols." *Journal of Educational Psychology* 39: 27-34; January 1948.

53. MALTER, MORTON S. "The Ability of Children To Read Cross-Sections." *Journal of Educational Psychology* 38: 157-66; March 1947.
54. MASON, JOHN M. "A Pre-Test and Post-Test Study in Biological Science." *Journal of Educational Research* 42: 228-33; November 1948.
55. MILLER, LEO R. "Some Effects of Radio-Listening on the Efficiency of Reading-Type Study Activities." *Journal of Educational Psychology* 38: 105-18; February 1947.
56. MITCHELL, ADELLE H. "The Effect of Radio Programs on Silent Reading Achievement of Ninety-One Sixth Grade Students." *Journal of Educational Research* 42: 460-70; February 1949.
57. NAHM, HELEN. "Mental Hygiene Knowledge of Senior Students in Schools of Nursing." *Journal of Educational Research* 41: 193-203; November 1947.
58. OLANDER, HERBERT T.; VAN WAGENEN, MARVIN J.; and BISHOP, HELEN M. "Predicting Arithmetic Achievement." *Journal of Educational Research* 43: 66-73; September 1949.
59. OTTO, HENRY J. "Comparison of Selected Organization and Administration Practices in 286 Public Elementary Schools and Forty-Six Campus Demonstration Schools." *Journal of Educational Research* 41: 81-87; October 1947.
60. PIERSON, GEORGE A. "School Marks and Success in Engineering." *Educational and Psychological Measurement* 7: 612-17; Autumn 1947.
61. RAMHARTER, HAZEL K., and JOHNSON, HARRY C. "Method of Attack Used by 'Good' and 'Poor' Achievers in Attempting To Correct Errors in Six Types of Subtraction Involving Fractions." *Journal of Educational Research* 42: 586-97; April 1947.
62. RINSLAND, HENRY D. "Form for Briefing and Evaluating Standardized Tests." *Journal of Educational Research* 42: 371-75; January 1949.
63. ROBINSON, FRANCIS P. "Effect of Language Style on Reading Performance." *Journal of Educational Psychology* 38: 149-56; March 1947.
64. RUSSELL, DAVID H. "Spelling Ability in Relation to Reading and Vocabulary Achievements." *Elementary English Review* 23: 32-37; January 1946.
65. RYAN, SISTER MARY MIRIAM. "Intercorrelations in Native Capacity and Accomplishments in Mathematics Courses." *School and Society* 70: 183-84; September 17, 1949.
66. SCATES, DOUGLAS E. "Fifty Years of Objective Measurement and Research in Education." *Journal of Educational Research* 41: 241-64; December 1947.
67. SEYMOUR, HOWARD C. "Counselor and the High School Testing Program." *Educational and Psychological Measurement* 6: 73-79; Spring 1946.
68. SIMPSON, RAY H. "The Critical Interpretation of Test Results in a School System." *Educational and Psychological Measurement* 7: 61-66; Spring 1947.
69. SPERZEL, EDITH Z. "The Effect of Comic Books on Vocabulary Growth and Reading Comprehension." *Elementary English Review* 25: 190-13; February 1948.
70. SUPER, DONALD E.; BRAASCH, WILLIAM F.; and SHAY, JOSEPH B. "The Effect of Distractions on Test Results." *Journal of Educational Psychology* 38: 373-77; October 1947.
71. SWENSON, ESTHER J. "When Is Test Performance Good Enough?" *Elementary School Journal* 48: 115-20; November 1947.
72. TIEDEMAN, HERMAN R. "A Study in Retention of Classroom Learning." *Journal of Educational Research* 41: 516-31; March 1948.
73. TILTON, JOHN W. "An Experimental Effort To Change the Achievement Test Profile." *Journal of Experimental Education* 15: 318-23; June 1947.
74. TOWNSEND, AGATHA. "An Investigation of Certain Relationships of Spelling with Reading and Academic Aptitude." *Journal of Educational Research* 40: 465-71; February 1947.
75. TRAXLER, ARTHUR E. "Evaluation of Methods of Individual Appraisal in Counseling." *Occupations* 26: 85-91; November 1947.
76. TRAXLER, ARTHUR E. "Measurement in the Field of Reading." *English Journal* 38: 143-49; March 1949.
77. TRAXLER, ARTHUR E. "A Testing Program for Mathematics at the Secondary School Level." *Mathematics Teacher* 39: 303-13; November 1946.
78. TRIGGS, FRANCES O., chairman. "Description of the Purposes and Functions of the Diagnostic Reading Tests." *Educational and Psychological Measurement* 8: 3-14; Spring 1948.

79. TROYER, MAURICE E. "An Attempt To Improve the Comprehensive Examination at the Masters Level." *Educational and Psychological Measurement* 6: 235-47; Summer 1946.
80. TROYER, MAURICE E. "An Evaluation Service Center." *Journal of Higher Education* 20: 152-55; March 1949.
81. VALLANCE, THEODORE R. "A Comparison of Essay and Objective Examinations as Learning Experiences." *Journal of Educational Research* 41: 279-88; December 1947.
82. VOTAW, DAVID F. "A Comparison of Test Scores of Entering College Freshmen as Instruments for Predicting Subsequent Scholarship." *Journal of Educational Research* 40: 215-18; November 1946.
83. WATSON, EUGENE P. "Final Examination Practices." *Journal of Higher Education* 17: 151-52; March 1946.
84. WATTS, PHYLLIS W. "Application of Clinical Diagnostic Techniques in the Classroom Situation for the Improvement of Reading at the College Level." *Journal of Educational Research* 42: 513-24; March 1949.
85. WEINLAND, JAMES D. "Attempt To Measure Cribbing in Objective Measurement." *Journal of Educational Sociology* 21: 97-100; October 1947.
86. WELBORN, ERNEST L. "The Scholarship of Veterans Attending a Teachers College." *Journal of Educational Research* 40: 209-14; November 1946.
87. WIEMAN, MARION. "Terminal Integrating Examinations." *Journal of Higher Education* 20: 161, 167-68; March 1949.
88. WRIGHTSTONE, J. WAYNE. "Frontiers in Educational Research in the Measurement of Aptitudes and Achievement." *Journal of Educational Research* 40: 389-96; January 1947.

CHAPTER VII

Educational Achievement Measures in Scholarship and Award Contests

IRVING LORGE AND ROSE KUSHNER

THIS is the first time an integrated treatment of the use of tests and other devices for the selection of gifted students in scholarship and award contests has appeared in the REVIEW. Accordingly, the time covered is somewhat greater than the usual three-year span. Consideration was given to such achievement tests as those made for the magazines *Time* and *Newsweek*, but this topic was excluded because no published research was found.

Methods for Selection of Award Winners

Findley (14) reported the inauguration of a comprehensive achievement test for the awarding of New York State scholarships. Prior to 1939 the New York State Regents scholarships were awarded on the basis of marks on the Regents subjectmatter tests. Beginning with 1939 an experiment was launched to determine whether a comprehensive examination based on the core-curriculum in secondary schools would be an acceptable substitute for subjectmatter averages. The comprehensive examination included 483 objective items from English, social studies, science, mathematics, art, and music, and two 200- to 300-word essays. Against a criterion of first- and second-year college grades, the correlation of comprehensive examination scores was not significantly different from that of the Regents averages. In general, the correlation between achievement test scores and grades was significantly higher than that between psychological examination scores and grades. Correlations of marks on the essay part of the achievement examination with first-year college grades for students in thirteen different colleges averaged about .25. The University of the state of New York reported (16) that the Regents program currently gives some weight to the results of the Regents tests for a chosen three-year sequence in a subjectmatter field. The same source also reported (18) the use of tests in awarding scholarships to veterans, war orphans, and applicants for medical schools.

Edgerton and Britt (1, 3, 4, 8) were associated with the annual Science Talent Search (for the awarding of four-year college scholarships) at its inception. Their method has been that of "successive hurdles," which involves, successively, a science aptitude examination, an evaluation of high-school scholarship record, recommendations of high-school teachers, evaluation of an essay, a personal interview, and some other measures of attitude and achievement. Hoffman (15) attacked the validity of the

"successive hurdles" method. Edgerton and Britt (2) recognized that the achievement examination favored the high-school students who were eclectic in choosing courses in science and in mathematics as opposed to students who tended to specialize.

The Pepsi-Cola program (19) also used the method of "selective hurdles." The procedure began with the nomination by their classmates of the 5 percent of high-school seniors "most likely to succeed." The nominees then took the scholarship examination, which was Program I of the College Entrance Examination Board. The examination consisted of a verbal and a mathematical section. The total score was obtained by adding one-fifth of the mathematics score to the verbal score. In 1947 (21) the program was expanded to include the awarding of graduate fellowships. In 1948 Stalnaker (20) reported on the practice of awarding scholarships to the best of the southern Negro high-school seniors. The method involved the cooperation of the segregated high schools of the southern states. The Negro high-school seniors nominated the 5 percent most likely "to make an important contribution to human progress." Each school, however, could send two representatives regardless of size of school, altho the upper limit was 5 percent of the senior class. These candidates took a preliminary selection test from which the eight highest scoring Negro pupils in each state took the scholarship examination. In each state the highest scoring Negro student was awarded a full four-year scholarship regardless of the level of his score.

Group Differences

Stalnaker (22) reported racial differences in scores earned by finalists in the Pepsi-Cola scholarship contests. The average scores for the four-year scholarship winners were 762 on the verbal section and 705 on the mathematics section. For the eighteen Negro winners (20), the corresponding scores were 488 and 436. It was pointed out that this difference may represent inequality of opportunity not only within the schools which the Negroes attended but also in the stimulations their environments afforded.

Most of the research studies noted an inequality in the results of girls as opposed to boys. The report of the Regents scholarship examination (16) indicated that the scholarships went to boys in the ratio of approximately three to one despite the facts that the number of girl high-school students is equal to that of the boys and that the ratio of boy to girl applicants is approximately three to two. In 1948 (17) a study of the examination showed no significant sex differences on Part I (social studies, art, music, literature) or on Part II (essay). Significant differences, however, were found on Part III (vocabulary, spelling, English grammar, mathematics, science, health, homemaking, and technical information). The boys did significantly better than the girls on the mathematics, science, and technical sections of the test. The girls did better on the other parts of the examination. Part III, however, was heavily weighted in favor of a

mathematics-science sequence. Over half of that part of the examination was based on these two subjects. Some people believe that giving credit for a three-year sequence would overcome such specialization. It should be pointed out, however, that most of the scholarship award winners presented a mathematics or a science sequence—89 percent of the boy and 53 percent of the girl winners. Edgerton and Britt (6) also showed test results in favor of the boys. In each of the science aptitude examinations the boys did significantly better than the girls.

Reliabilities

Edgerton and Britt (7) discussed the several aspects of the hurdles in the Fourth Annual Science Talent Search. They estimated the reliability of the first two parts of the achievement test at about .75 and the average intercorrelations among the three parts at .58. The achievement test had three different parts based on objective-type items. The parts were (a) multiple-choice aptitude in various areas of science, (b) paragraph reading in science, and (c) general science information and items testing the ability to make inferences. In another study, these authors (10) indicated that the reliability of anecdotal material ranged from .50 to .85 and that the recommendations by teachers correlated only .20 with the aptitude examination itself.

Value of Essays

The value of essays in the selection of candidates was investigated in several studies (7, 16, 17). The reliability of essay ratings of the Science Talent Search examination (7) based on average intercorrelations of three essay raters was .63 for boys' essays and .75 for girls' essays. Using contest standing as the criterion, the correlation for the rating of the essay was .43 and the corresponding correlation for the achievement test was about .48. Both aspects of the examination were considered equally effective in the selection process. The essay part of the New York State Regents Scholarship examination affected the ranking of 91 candidates out of 827, and hence was considered to have high discriminatory value (17).

Follow-Up Validity

The Science Talent Search published the results of a few studies on follow-up. They indicated (11) that the scholarship winners tended to select chemistry and physics for their major field significantly more often than did the runners-up. As might be expected, the winners got higher grades and more elections to honorary societies than did the runners-up of the general run of students. A significant byproduct of the follow-up was the demonstration that the winners tended to respond to mailed ques-

tionnaires (12) more frequently than did the other contestants. Edgerton and Britt (5, 9) pointed out that the number of Science Talent Search winners was not proportional to the number of high-school seniors in each state. For instance, some states high in educational and economic indices took a disproportionately higher number of winners. Their evidence suggested (13) that the intellectually abler members of the population are also physically superior.

Summary

It is significant that all the research studies reported that more boys than girls were awarded scholarships. In the Science Talent Search the well-known tendency of boys to favor science and mathematics should, of course, show in better scores for the boys. In the Pepsi-Cola and New York State scholarship programs, however, the implication is that the examination was made to select the intellectually gifted as such, regardless of the areas of specialization. Since more boys than girls are selected, research is needed to discover whether a bias exists in the examinations and in the selection procedures in favor of students who have taken science and mathematics at the expense of students who have had courses in the humanities. It is certainly implicit that intellectual aptitude can produce in fields other than in science and mathematics. It is hoped that the next few years will show more work on follow-up of the scholarship winners and the runners-up. The students selected in these various programs undoubtedly come from the upper 3 percent of the high-school population, and their success should prove a guidepost for all selection programs.

All these programs tended to emphasize science and mathematics, either directly or indirectly. It is certainly encouraging that one state and two outside agencies have tried to discover the intellectually gifted. It is hoped that future studies will give evidence about the genuine validity of the various procedures used. The major problem is to discover whether the intellectually gifted choose science and mathematics or whether those of the gifted who study other subjectmatter areas are penalized by the examinations.

Bibliography

1. EDGERTON, HAROLD A., and BRITT, STEUART H. "The First Annual Science Talent Search." *American Scientist* 31: 55-68; Winter 1943.
2. EDGERTON, HAROLD A., and BRITT, STEUART H. "Further Remarks Regarding the Science Talent Search." *American Scientist* 31: 263-65; July 1943.
3. EDGERTON, HAROLD A., and BRITT, STEUART H. "How Science Talent Search Winners Are Chosen." *Science and the Future*. Essays and Reports of the Second Annual Science Talent Search for the Westinghouse Scholarships. Washington, D. C.: Science Service, Incorporated, 1943.
4. EDGERTON, HAROLD A., and BRITT, STEUART H. "The Science Talent Search." *Occupations* 22: 177-80; December 1943.

5. EDGERTON, HAROLD A., and BRITT, STEUART H. "The Science Talent Search in Relation to Educational and Economic Indices." *School and Society* 63: 172-75; March 9, 1946.
6. EDGERTON, HAROLD A., and BRITT, STEUART H. "Sex Differences in the Science Talent Test." *Science* 100: 192-93; September 1, 1944.
7. EDGERTON, HAROLD A., and BRITT, STEUART H. "Technical Aspects of the Fourth Annual Science Talent Search." *Educational and Psychological Measurement* 7: 3-21; Spring 1947.
8. EDGERTON, HAROLD A., and BRITT, STEUART H. "The Third Annual Science Talent Search." *Science* 99: 319-20; April 21, 1944.
9. EDGERTON, HAROLD A.; BRITT, STEUART H.; and DAVIS, HELEN M. "Is Your State Discovering Its Science Talent?" *Science Education* 28: 228-31; October 1944.
10. EDGERTON, HAROLD A.; BRITT, STEUART H.; and LEMMON, WILLIAM B. "Reliability of Anecdotal Material in the First Annual Science Talent Search." *Journal of Applied Psychology* 31: 413-24; August 1947.
11. EDGERTON, HAROLD A.; BRITT, STEUART H.; and NORMAN, RALPH D. "Later Achievements of Male Contestants in the First Annual Science Talent Search." *American Scientist* 36: 403-14; Summer 1948.
12. EDGERTON, HAROLD A.; BRITT, STEUART H.; and NORMAN, RALPH D. "Objective Differences Among Various Types of Respondents to a Mailed Questionnaire." *American Sociological Review* 12: 435-44; August 1947.
13. EDGERTON, HAROLD A.; BRITT, STEUART H.; and NORMAN, RALPH D. "Physical Differences Between Ranking and Non-Ranking Contestants in the First Annual Science Talent Search." *American Journal of Physical Anthropology* 5: 435-52; December 1947.
14. FINDLEY, WARREN G. "The Validity of a Comprehensive Examination for Scholarship Awards in New York State." *Journal of Experimental Education* 11: 250-56; March 1943.
15. HOFFMAN, BANESH. "Some Remarks Concerning the 'First Annual Talent Search.'" *American Scientist* 31: 251-62; July 1943.
16. NEW YORK (STATE) UNIVERSITY. *Report on Regents Scholarship Examination—1947 Series*. Albany: the University, December 1947. 5 p. (Mimeo.)
17. NEW YORK (STATE) UNIVERSITY. *Report on Regents Scholarship Examination—1948 Series*. Albany: the University, December 1948. 20 p. (Mimeo.)
18. NEW YORK (STATE) UNIVERSITY. *State War Service Scholarships for Veterans: Law, Regulations, and Instruction to Candidates*. Handbook 46. Albany: the University, June 1949.
19. STALNAKER, JOHN M. "A Commercially Financed Scholarship Program." *School and Society* 62: 60-61; July 28, 1945.
20. STALNAKER, JOHN M. "Identification of the Best Southern Negro High-School Seniors." *Scientific Monthly* 67: 237-39; September 1948.
21. STALNAKER, JOHN M. "Pepsi-Cola Scholarship Board Activities." *School and Society* 66: 396-97; November 22, 1947.
22. STALNAKER, JOHN M. "Results of Pepsi-Cola Finalists on the College Board Test." *School and Society* 68: 110-11; August 14, 1948.

CHAPTER VIII

Measurement of Educational Achievement in Nonschool Educational Agencies and in Industry

J. RAYMOND GERBERICH AND JAMES M. BURKE

SEVERAL types of achievement testing not treated elsewhere in this issue and not treated in an integrated manner in any recent issues of the REVIEW are the programs carried on by nonschool educational agencies and the testing conducted in industrial and other nonschool settings. In some respects and for some programs the applications of achievement tests of this section parallel the application of aptitude tests discussed by Stuit in Chapter III of this issue of the REVIEW.

General Educational Development

In the December 1945 issue of the REVIEW Gerberich discussed the inception and early development of the *U. S. Armed Forces Institute Tests of Educational Development*, and Schrader and Conrad in the December 1948 REVIEW brought research on these tests and on large-scale testing programs up to date. The present account, therefore, deals primarily with such testing programs during the past year.

Findley and Andregg (14) obtained data from the administration of the *USAFI Tests of General Educational Development* to more than 1000 junior Air Force officers in the Air Tactical School at Maxwell Field, Alabama. These data were statistically analyzed to determine the reliability of the tests, the correlation of test scores with amount of education and with school achievement, the comparability of test intercorrelations with intercorrelations among West Point marks in the same school subjects, the values of the tests in differential diagnosis, their factorial composition, and the average item validities. The authors concluded from their findings that these tests possess practical validity, in addition to face validity, which could be useful in selecting Air Force officers to study at civilian colleges and universities. They also reported that more than two million veterans and other persons had taken the *GED* tests by June 1947. Donohue (10) established University of Michigan norms for the *USAFI Tests of General Educational Development* which were slightly higher than the norms offered by USAFI for Type I colleges.

Putnam (27) compared the scholastic achievement and number of withdrawals of a group of non-high-school graduates admitted on the basis of results on the tests of *General Educational Development* to the Vanport Extension Center with similar data for a group of students admitted to the center on the basis of their high-school graduation. He concluded that high-school graduation is not essential to successful scholastic achievement

in college and that a properly motivated student of average aptitude can, after three years of high school, compete successfully with high-school graduates. Batmale (6) studied the performance of 300 veterans on the *USAFI Tests of General Educational Development*. These veterans had left high school before graduation (the average veteran had left in low Grade XI) and had achieved high-school graduate status as a result of success on the tests. The veterans were found to compare favorably with the standardizing groups in all tests except the *Correctness and Effectiveness of Expression Test*, where they performed significantly lower.

Integrated Admissions Testing Programs

The three integrated examination programs designed to aid in the selection of undergraduate college students, graduate college students, and school teachers are, respectively, the tests of the College Entrance Examination Board, the *Graduate Record Examination*, and the *National Teacher Examinations*. The activities of all three programs are now directed by the Educational Testing Service.

The forty-seventh (9) and forty-eighth (8) annual reports of the directors summarized the testing and research activities of the College Entrance Examination Board during 1947 and 1948 and furnished validity and other data about the tests. Riegel (28) reported a validity study by Schrader and Fredericksen on the values of College Board mathematics tests for predicting first-semester marks in five engineering colleges. Schultz and Plumlee (32) studied the comparability of the three types of mathematics tests offered by the College Board. Low but significant correlations were found in a study by Peixotto (26) between the *English Essay Test* and the *Verbal Scholastic Aptitude Test* and between the *English Essay Test* and the *Cooperative Reading Comprehension Test*, C2. Dyer (11) found correlations ranging from .64 to .94 between the College Entrance Examination Board language test scores and final marks in corresponding elementary language courses.

Vaughn (38) and Ryans (31) discussed respectively the *Graduate Record Examination* and the *National Teacher Examinations* in terms of background and function. Speer (33) pointed out the uses of the *Graduate Record Examination* in selecting graduate engineering students, and a new norms bulletin reporting the achievement of senior students on the advanced test (12) was issued. Ryans (30) presented and analyzed results for the February administration of the *NTE*.

Evaluating and Selecting Personnel

The use of achievement measures in employee evaluation and selection is currently receiving attention. Thorndike (35) produced a volume dealing comprehensively not only with the development and appraisal of a per-

sonnel testing program but also with the problems involved in the effective administration of such a program. A War Department Technical Manual (37) explained the development of Army measurement procedures for use in personnel selection and management.

Measurement of achievement thru job information and work sample tests has been used both independently and in conjunction with aptitude measurement in the selection of personnel for jobs and in personnel evaluation through merit ratings and judgments of efficiency as bases for promotions.

Adkins (3) outlined procedures in construction and analysis of tests designed to predict job performance. Later Adkins (2) presented a broader and more detailed exposition of procedures for construction and analysis of both paper-and-pencil and work-sample tests based on job analysis. Moore (25) discussed the use and extent of tests currently in use in industry. Flanagan (15) called for a scientific approach to the problems of evaluating personnel based on definitions of job requirements; observation of work performance; and summarizing, interpreting, and using the data obtained.

Much of the research in the use of such tests with personnel in the technical and skilled labor fields comes from the armed forces. Goodman (17) presented a validity study of written job-knowledge tests in the army food service field. Henneman (20) discussed the need for placing greater emphasis on determining critical standards. He described four sources of proficiency data for measuring military technical specialists: technical information tests, practical performance tests, work production records, and subjective assessments of worker proficiency. Klein (21) attempted to evaluate self-appraisal of test performance in dealing with an aircrew selection problem. He concluded that cadets who overestimate their performance are more prone to failure in flying training than those who underestimate their performance.

Rosenberger (29) presented the development and use of the performance tests of the United States Bureau of Prisons for selection of six types of occupations needed in a penal institution. Bean (7) offered various techniques for constructing an English usage test for use with prospective clerical personnel. Stewart (34) attempted to determine from a study of the results of a civil service examination for clerk-typists the factors which figured in the failure of 449 out of 551 applicants. She identified deficiencies in spelling skills, speed in filing and checking, and typing speed and accuracy. Ekberg (13) devised a mechanical performance test to measure efficiency in certain mechanical jobs, presenting it as reasonable and valid if used as only one in a battery of tests. Lefever, Van Boven, and Banarer (23) presented evidence on the validity of job information tests constructed to meet the needs of certain training programs related to the repair and maintenance of airplanes. These same authors (22) followed this with a study of job information tests for airplane mechanics and warehousemen in the attempt to discover the effect of age and amount of schooling on

success in the test. They found that the age and education factors favored but slightly the workers between the ages of twenty and fifty.

The use of job performance technics has also received attention in the more complex fields of administration, supervision, and leadership. The Office of Strategic Services (36) made a thoroughgoing attempt to select men and women of initiative, courage, and responsible leadership, leaning heavily on the use of situational tests. Mandell (24) reported on an examination program for the selection and promotion of foremen in various field installations of the Navy department. Abt (1) described a test battery for the selection of editors of technical magazines. Part of this battery was a series of seven work-sample tests of usual editorial work. The results showed significant differences between "good" and "bad" criterion groups.

Miscellaneous Studies

Gilbert and Wrightstone (16) evaluated psychological and educational outcomes of an extension of schooling by means of sending pupils to camp for three weeks during the school term. The experimental pupils, those who went to camp, were matched with control pupils who did not go. The evaluation showed definitely greater gains by the camp group than by the control group in content areas and range of interests. The Merit System Service of the American Public Health Association (5) demonstrated the types of tests used in its program and gave detailed item analysis data to illustrate its procedures in test validation.

The Air University at Maxwell Field, Alabama, presents an unusual project in higher education in which instruction is carried on by subject-matter experts with little or no background in professional education. Educators, however, are used in an advisory capacity. A manual for test construction and evaluation (4) was issued by the Educational Services Division. Several reports of the use of evaluation procedures in the improvement of instruction have also appeared in the literature. Greene and Findley (19) outlined a program for improvement of instruction in higher education by means of evaluative procedures previously elaborated by Greene (18) in an Air University study based on three areas: improving the measurement of intangibles; improving objective test items; and utilizing student evaluations.

Bibliography

1. ABT, LAWRENCE E. "A Test Battery for Selecting Technical Magazine Editors." *Personnel Psychology* 1: 75-91; Spring 1949.
2. ADKINS, DOROTHY C. *Construction and Analysis of Achievement Tests*. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947. 292 p.
3. ADKINS, DOROTHY C. "Construction and Analysis of Written Tests for Predicting Job Performance." *Educational and Psychological Measurement* 6: 195-211; Summer 1946.

4. AIR UNIVERSITY, EDUCATIONAL SERVICES DIVISION, TEST CONSTRUCTION UNIT. *Test Construction and Interpretation*. Bulletin No. 1. Maxwell Field, Alabama: Air University, August 1946. 30 p.
5. AMERICAN PUBLIC HEALTH ASSOCIATION, MERIT SYSTEM SERVICE. *A Report on "Take a Test."* New York: the Association, 1949. 44 p. (Mimeo.)
6. BATMALE, LOUIS F. "Veterans' High-School Graduation by Examination." *School Review* 56: 229-35; April 1948.
7. BEAN, KENNETH L. "The Development of an English Usage Test for Clerks, Typists, and Stenographers." *Educational and Psychological Measurement* 6: 331-39; Autumn 1946.
8. COLLEGE ENTRANCE EXAMINATION BOARD. "The Board and Its Tests." *Forty-Eighth Annual Report of the Director, 1948*. New York: the Board, 1948. p. 9-49.
9. COLLEGE ENTRANCE EXAMINATION BOARD. "Tests for College Entrance." *Forty-Seventh Annual Report of the Director, 1947*. New York: the Board, 1947. p. 7-45.
10. DONAHUE, WILMA F. "University of Michigan Norms for the United States Armed Forces Institute Tests of General Educational Development." *Educational and Psychological Measurement* 6: 261-64; Summer 1946.
11. DYER, HENRY S. "Some Observations on the College Board Language Tests." *Educational and Psychological Measurement* 8: 593-602; Winter 1948.
12. EDUCATIONAL TESTING SERVICE. *Graduate Record Examination, Norms Bulletin: The Performance of Senior Students on the Advanced Tests*. Bulletin No. 3. New York: Educational Testing Service, 1948. 7 p.
13. EKBERG, DAVID L. "A Study in Tool Usage." *Educational and Psychological Measurement* 7: 421-27; Autumn 1947.
14. FINDLEY, WARREN G., and ANDREGG, NEAL B. "A Statistical Critique of the USAFI Tests of General Educational Development." *Psychometrika* 14: 47-60; March 1949.
15. FLANAGAN, JOHN C. "A New Approach to Evaluating Personnel." *Personnel* 26: 35-42; July 1949.
16. GILBERT, HARRY B., and WRIGHTSTONE, J. WAYNE. "Evaluation of Psychological and Educational Changes in a Three Week School-Camp Experience." *American Psychologist* 3: 297; July 1948.
17. GOODMAN, CHARLES H. "Validation of Job Proficiency Tests for the Army Food Service Field." *American Psychologist* 3: 305-306; July 1948.
18. GREENE, JAMES E. "The Evaluation of Instruction in Air University." *Journal of Psychology* 25: 279-97; April 1948.
19. GREENE, JAMES E., and FINDLEY, WARREN G. "Evaluative Procedures for the Improvement of Instruction." *Educational Record* 30: 33-44; January 1949.
20. HENNEMAN, RICHARD H. "The Development of Proficiency Measures for Military Technical Specialists." *American Psychologist* 3: 304-305; July 1948.
21. KLEIN, GEORGE S. "Self-Appraisal of Test Performance as a Vocational Selection Device." *Educational and Psychological Measurement* 8: 69-84; Spring 1948.
22. LEFEVER, D. WELTY; VAN BOVEN, ALICE; and BANARER, JOSEPH. "Relation of Test Scores to Age and Education for Adult Workers." *Educational and Psychological Measurement* 6: 351-60; Autumn 1946.
23. LEFEVER, D. WELTY; VAN BOVEN, ALICE; and BANARER, JOSEPH. "Validation Studies in Job Information Tests." *Educational and Psychological Measurement* 6: 223-33; Summer 1946.
24. MANDELL, MILTON M. "The Selection of Foremen." *Educational and Psychological Measurement* 7: 385-97; Autumn 1947.
25. MOORE, HERBERT. "Current Tests in Industry." *American Psychologist* 2: 321; August 1947.
26. PEIXOTTO, HELEN E. "The Relationship of College Board Examination Scores and Reading Scores for College Freshmen." *Journal of Applied Psychology* 30: 406-11; August 1946.
27. PUTNAM, PHIL H. "Scholastic Achievement of GED Students at the Vanport Extension Center." *School and Society* 66: 161-63; August 30, 1947.
28. RIEGEL, ELEANOR J. "Evolution of Math Tests." *College Board Review* 1: 95-98, 102; May 1949.
29. ROSENBERGER, HOMER T. "Testing Occupational Training and Experience." *Educational and Psychological Measurement* 8: 101-15; Spring 1948.

30. RYANS, DAVID G. "The 1948 National Teacher Examinations." *Journal of Experimental Education* 17: 1-25; September 1948.
31. RYANS, DAVID G. "The Use of the National Teacher Examinations in Colleges and Universities." *Journal of Educational Research* 42: 678-89; May 1949.
32. SCHULTZ, DOUGLAS, and PLUMLEE, LYNNETTE B. "Scores on Board Mathematics Examinations Compared." *College Board Review* 1: 81, 84-87; March 1949.
33. SPEER, GEORGE S. "The Use of the Graduate Record Examination in the Selection of Graduate Engineering Students." *Journal of Engineering Education* 37: 313-18; December 1946.
34. STEWART, JANE C. "A Study of the Failures in a Los Angeles Civil Service Clerk-Typist Examination." *National Business Education Quarterly* 16: 55-60; October 1947.
35. THORNDIKE, ROBERT L. *Personnel Selection: Tests and Measurement Techniques*. New York: John Wiley and Sons, 1949. 358 p.
36. U. S. OFFICE OF STRATEGIC SERVICES, ASSESSMENT STAFF. *Assessment of Men*. New York: Rinehart and Company, 1948. 541 p.
37. U. S. WAR DEPARTMENT. *Personnel Classification Tests*. War Department Technical Manual, TM12-260. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1946. 90 p.
38. VAUGHN, KENNETH W. "The Graduate Record Examinations." *Educational and Psychological Measurement* 7: 745-56; Winter 1947.

CHAPTER IX

Construction and Validation of Educational Tests

ROBERT L. EBEL

Previous Reviews, Collected Papers, and Books

RESEARCH in the field of test construction and validation was reviewed most recently by Schrader and Conrad in the December 1948 issue of the REVIEW. The February 1947 issue included a chapter by Travers which dealt with item and test characteristics, reliability, validity, and factor analysis, among other matters. Meder and Eagle reviewed measurement in mathematics and science in the October 1948 issue.

Swineford and Holzinger (59) continued their useful annual compilations of references and statistics, the theory of test construction, and factor analysis, which appear in the *School Review*. New tests are reviewed periodically in the *Journal of Consulting Psychology*.

Each summer the *American Psychologist* (2, 3, 4) has published a program of the annual meeting of the American Psychological Association which includes abstracts of papers dealing with test construction and validation. Good's annual summaries (23) of "Doctoral Dissertations Under Way in Education" in the *Phi Delta Kappan* have listed twenty-five or more studies in the field of test construction and validation during the past three years. In each case the institution at which the research is being done and the sponsoring professor are listed. No doubt many of these studies have been completed at this writing.

A bulletin by Darley and others (13) dealt with the use of tests in college. Chapter 7 of this bulletin discussed fundamental principles in test construction. Research related to the construction and validation of educational tests was reported in *Growing Points in Educational Research* (1). Here Spencer reported on testing in arithmetic, Stegeman on spelling, Beck on language arts, and Walker on item analysis. Papers read at the annual Invitational Conference on Testing Problems, sponsored by the Educational Testing Service, were reported in three annual bulletins entitled respectively, *National Projects in Educational Measurement* (68), *Exploring Individual Differences* (11), and *Validity, Norms and the Verbal Factor* (20). Of special interest is the article by Lindquist (41), which described the projected manual on educational measurement. Donahue, Coombs, and Travers (18) published the papers read during the Guidance Conference on the Measurement of Student Adjustment and Achievement at Ann Arbor, Michigan, in June 1947.

More attention has been given in recent years to the publication of practical manuals on test construction. Weitzman and McNamara (75)

produced a book providing suggestions for and illustrations of good test-construction practices on a practical, nontechnical level. The Adjutant General's office (66) likewise prepared a simple, practical manual on test construction.

An event of great importance to users of educational tests was the publication in 1949 of Buros' *Third Mental Measurements Yearbook* (10). In this 1047-page volume, 705 educational tests or test batteries are listed and most are reviewed in detail by competent critics. In addition there are lists and reviews of more than 500 books dealing with tests and testing.

Hortatory Discussions

Literature on testing in education has always been replete with articles intended to direct, stimulate, or retard the construction and use of tests. Some of these articles have been written more on the basis of bias and misinformation than on the basis of experience, insight, and judgment. Others, pitched at a very elementary level, have over-simplified fundamental problems. Many of the articles, however, have made valuable contributions.

General suggestions for test construction were included in an article by Armstrong (5). Suggestions were offered for test construction in the foreign languages by Coutant (12), in arithmetic by Spitzer (58), in science by Weaver (74), and in vocational subjects by Beckley and Smith (7).

Thurstone (62) made a strong plea for truly cooperative efforts in the production of tests. If his arguments could move responsible organizations in various teaching fields to adopt the plan suggested, a general improvement in the quality of typical course examinations in that field would be almost certain to follow. Wall (73) also suggested cooperative testing efforts among physics teachers.

Sims (55), in a provocative article, stated and criticized five assumptions which he found underlying current achievement testing. It is not difficult to agree with Sims that motives, as well as ability, are important, that the interpretation of scores need not be based exclusively on norms, and that initial-terminal testing does not permit highly accurate projection of future status. But criticism of a measure of growth in one area for failing to consider possible offsetting deterioration in another area is difficult to accept. The suggestions of Sims that educators adopt testing procedures consistent with "the implications of insightful learning or of organismic or field psychology" is appropriate, but it is difficult to conceive of procedures suitable to accomplish this purpose.

Thorndike (61) in discussing the future of the measurement of abilities identified objectivity, adequacy, and purity as the basic elements of quality in tests. The difficulties with respect to the last two, and possible means of overcoming these difficulties, were discussed clearly.

Specific Problems in Test Construction

Bartlett (6) discussed the constituents of human skill, with some references to experimental data and some indications of further lines of research. Seashore and Bennett (53) used a standard set of dictated letters on records in their test of stenography. Careful attention was given to such problems as rate of presentation and method of scoring errors. The test possesses work-sample validity and shows satisfactory correlation with efficiency ratings. Kenneally (37) constructed a test of study skills whose value was defended in part on the basis of increase in average score from grade to grade. One basis used for determining the validity of the test was correlation between scores on it and on the *Iowa Silent Reading Test*. Both of these procedures are open to some question.

Dayshaw (17) reviewed comprehensively the history and problems of the measurement of interest, discussing the types of tests used, the scoring and scaling methods employed, and the relationships discovered. He suggested new approaches and reported an extensive list of references. An achievement test was used by Peel (50) as a disguised test of interest. The items were divided into groups within each of which both practical and academic abilities were represented. The examinees, permitted to answer only a fraction of the items in each group, revealed interests by the choices made. It was found that rather sharp distinctions between practically minded and academically minded examinees were possible. Turner (64) reported a scale of altruism based on informers' reports on subjects' responses to certain situations. A scale for measuring self-insight was prepared by Gross (26). The scale consists of thirty-seven statements similar to this, "I have always appreciated frank criticism of my faults." The statements in the scale are either true but not flattering for a majority of the people, or false but flattering for a majority of the people. The examinee's self-insight is determined by the extent to which he accepts the true, unflattering statements and rejects the false, flattering statements.

In a carefully executed study, Villarreal (71) constructed a test of aural comprehension of English for native speakers of Spanish. The test is based on oral passages and attempts to control possibly interfering variables such as hearing loss, distortion, volume, room conditions, rate, and regional differences. The test was validated against self ratings and ratings by acquaintances. Griffin (24) reported on a diagnostic test for adults of limited reading ability.

The problem of measuring speech hearing was analyzed by Irwin (33). On the basis of his analysis, three tests measuring word meaning and perception of speech sounds were constructed. The tests consist essentially of sentences which can be completed by using words having different sounds and meanings. The complete test sentences are presented orally to the examinee by means of phonograph recordings. The examinee's task is to identify on his answer sheet which of several words was used in the

sentence. Irwin's report of the study purports to show that the sample of subjects used was "representative of the population," but fails to identify the population.

Heider (29) described the construction of the language usage test for the deaf, in which item selection was based on a study of errors made by deaf children in their letters home. Blattner (8) investigated various methods of testing pronunciation but did not find any which was completely satisfactory.

In the science field, Friedenbergr (22) reported attempts to measure student insight into the basic structure of biological science. Dunning (19) explained the construction and evaluation of a test of scientific thinking. A test of critical judgment by Ullsvik (65) took as its point of departure a definition of critical judgment.

Grigg (25) prepared a thirty-item farm knowledge test which discriminated clearly between people with rural and urban backgrounds. It is not surprising that a test which discriminates at this level did not correlate highly with years of farm experience.

Item Types

Publications in the period covered by this REVIEW reveal very little in the way of new item forms. No doubt this is wholly commendable. There is certainly less need for new item forms than for improved use of existing forms. It is regrettable, however, that no really comprehensive descriptive catalog of test items designed to measure diverse educational outcomes has yet been prepared.

Testing in the armed forces has led to extensive development of test items based on pictures. Such items are especially well suited to problems dealing with equipment and terrain, but they deserve wider usage in civilian education. The Department of the Army (67) recently published a comprehensive and practical manual dealing with the construction and use of test items based on pictures.

Montgomery (45) described the construction, administration, and scoring of test items which require things to be arranged in order on the basis of some criterion. The scoring of this test is based on the displacement of each element from its proper position in the series. While this scoring may be quite valid for tests which emphasize quantitative ranking, its validity is less evident where emphasis is placed on sequence rather than on rank.

Troyer and Angell (63) developed a scoring device which is applied to multiple-choice items but which appears to alter the function of those items considerably. The examinee indicates his responses by punching a hole in the answer sheet held in a special frame. When the correct response to an item has been punched, a red dot appears thru the hole. The examinee is directed to continue responding to each item until he has selected the

correct response. His score, an inverse measure of ability, is the total number of punches made. This device is recommended largely on the ground that it reduces the teacher's labor and increases the student's learning. Experimental data were presented to support the second of these contentions. Its possible effects on the reliability and validity of measurement remain to be determined. A study of this device by Jones and Sawyer (36) included a summary of pupil comments on its effects.

Item Analysis and Selection

One of the essential steps in the construction of a valid test is the analysis and selection of test items. Research in this area has been concerned chiefly with procedures for computing item analysis data, studies of the characteristics of the indices obtained and their relations to other variables and methods for item selection on the basis of the analytic data.

A review of methods of item analysis was presented by Vernon (69) who, in the tradition of item analysts, suggested a new method for which certain advantages were claimed. The article was accompanied by an extensive bibliography. An index of item validity giving comparable values (a) for items of equal discriminating power at all levels of difficulty, and (b) for items having different numbers of responses was proposed by Johnson (35).

Walker (72) suggested application of the methods of sequential sampling to item analysis. The basic purpose of this method is to provide a test of the hypothesis that there is no relation between the criterion score and correctness of response to the item under consideration. Sequential sampling saves labor by limiting the sample size to approximately the minimum number of cases which will permit the rejection of the null hypothesis. This method does not yield directly an index of discrimination. Even tho it is probably less laborious to use than the biserial r when the necessary tables have been computed, it is far more laborious than other commonly used methods which yield quite satisfactory indices.

A study by Davis (16) based on his previously published paper and item analysis chart (15) confirmed other findings that indices of discrimination tend to be considerably less reliable than indices of difficulty. Wesman (76) studied the effect of restrictive time limits upon item-test correlation coefficients and found support for the logical conclusion that values obtained for items placed toward the end of a speeded test are almost worthless. Mensh (44) found no significant difference between responses to the same items when presented in short, medium, or long forms of a test. The use of the IBM Graphic Item Counter to obtain data efficiently in computing item test intercorrelations was described by Mount (46).

Kirkpatrick and Cureton (39) investigated the relation between the difficulty of vocabulary items and the frequency in popular usage of the

key word in each item. They made a useful distinction between vocabulary tests which measure range and those which measure precision. The correlation between frequency and difficulty was considerably higher for tests emphasizing range than tests emphasizing precision. Harris (28), studying the predictability of item difficulties, found low correlations between the difficulty indices of spelling test items and the mean scale values of the component words. An exception was found where the between-items variance in scale difficulty was large compared with the within-items variance.

The problem of using item analysis data most effectively has concerned a number of investigators. To avoid some of the questionable assumptions and practical difficulties involved in usual procedures, Loevinger (43) recommended test evaluation and item selection to maximize homogeneity of the test. She also set up criteria for an "adequate" system of score scaling which requires homogeneity in the test. Current methods of scaling do not appear to satisfy these criteria. Loevinger recognized the practical usefulness of heterogeneous, inadequately scaled tests and conceded that the validity of tests constructed according to her criteria remains to be established.

The relation of item analysis and selection to scaling was also discussed by Peel (51), who suggested a method purporting to yield what most experts in measurement have agreed is unobtainable—an absolute linear scale of ability. Brogden (9) considered the dependence of test validity on the distribution of item difficulties, the number of items, and the intercorrelations between them. Evidence on the relationship between the internal consistency of a test and the validity of the component items was presented by Owens (48). Lawshe and Mayer (40), using two methods of item analysis, studied the relation of item selection to test reliability.

Kinzer and Kinzer (38) analyzed the fifteen arithmetic test items in a chemistry placement examination. They found four items which in combination appeared as effective as the entire fifteen in predicting the final course mark. One weakness of this study is that the test of validity was not independent of the item selection, since both were based on the same set of responses. Because of this, purely chance relationships between the criterion and the responses were permitted to inflate the coefficient of predictive validity. It is practically certain that administration of the four selected items to a different but comparable group of students would yield scores showing much lower correlation with the criterion than that reported in this study.

Validation Studies

From both theoretical and practical points of view, the validation of educational tests continues to present troublesome problems. The research of this period reveals interesting suggestions for the solutions to some of the technical problems related to validity determination and commendable efforts to demonstrate validity thru correlation of test measures with

criterion measures. There has been, however, no comprehensive discussion of validity which will help to put validity coefficients for educational tests in a proper perspective. In particular, there is need for emphasis on the fact that judgments of face validity are involved inevitably in the validation process. If not applied to the test itself, then they must be applied in selecting criteria. It is quite apparent that in many instances the time and effort which have been devoted to the establishment of correlational validation on the basis of inadequate criteria accepted at face value would have been better spent in critical examination of the face value of the test itself.

The criterion has received considerable attention in recent research. Patterson (49) discussed some of the problems involved and recommended the use of refined ratings, in certain instances, as substitutes for more objective but less valid criteria. A reversal of the usual multiple correlation procedure was suggested by Hsu (32). To validate a single test Hsu suggested multiple criteria corrected for attenuation and weighted to maximize the multiple correlation. However, the "validity" of such an index of validity would be questionable. Peel (52) presented a solution of the problem of determining test weights which gives maximum prediction of a complex external criterion formed of a number of arbitrarily weighted assessments. A similar problem was considered by Thomson (60).

The interpretation of validity coefficients is facilitated by information concerning the reliability of the criterion measures. A generalized reliability formula, applicable where the number of available measures varies from one member of the group to another, was developed by Horst (31), who showed that several of the usual formulas constitute special cases of this more general formula.

The estimation of battery validity involves multiple correlation. Ordinarily the component tests of the battery are weighted without regard to the time they require. But the validity of a test is to a degree a function of its length, which, in turn, determines the time it requires. Long and Burr (43) developed a modification of the Wherry-Doolittle method of test selection so that tests are selected in order of their return in validity per unit of testing time. Considering the same problem, Horst (30) presented a method for apportioning time to tests in a battery with finite time limits in order to maximize predictive value.

A report by Fletcher and Hildreth (21) gives a broad and not too well-defined evaluation study of predictive validity. The invalidities of general reading tests were discussed by Shores (54), who suggested emphasis on currently overlooked factors and changes in testing procedure. Jarvis (34) illustrated in detail Hoyt's methods of test evaluation.

Spache (57) and White (77) used case studies to investigate validity. This method has much merit, but the labor involved greatly restricts its application. White also made use of a variety of special criteria in demonstrating the validity of an English test.

Factor Analysis

Factor analysis has assumed increasing importance in test construction and validation, while discussion of its functions and values continues. Guilford (27) considered the role of factor analysis in relation to the development of job tests and test batteries. He pointed to its value in suggesting new tests; in understanding existing tests, existing criteria, and the relation between tests and criteria; and in increasing the predictive value of test batteries. The contributions of factor analysis to test validation depend largely upon the inclusion of criteria as well as tests in the correlation matrix. This fact is sometimes overlooked. Factor analysis provides a convenient means for dealing with complex criteria. Guilford, an enthusiastic supporter of factor analysis, does not stress its limitations.

Most of the other research in this area involves reports of factor studies in various fields. Analyses of reading by Davis (14), a spatial ability by Smith (56), and of geometric ability by Murray (47) were reported. Vernon (70) executed a factor analysis of tests of practical ability which led him to the conclusion that a general practical ability exists but that it is amorphous and mainly an aggregate of nonsymbolic abilities. Vernon reported that except for the small dexterity component, paper-and-pencil tests are as useful as performance tests in assessing practical ability.

Bibliography

1. AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. *Growing Points in Educational Research*. 1949 Official Report. "The Study of Type Responses in Arithmetic Computation." (Spencer, Peter) p. 44-47; "Measurement and Control of Forgetting Rates in Spelling." (Stegeman, William H.) p. 47-50; "Evaluation in the Language Arts." (Beck, Roland L.) p. 152-55; "Item Selection by Sequential Sampling." (Walker, Helen M.) p. 181-87. Washington, D. C.: the Association, a department of the National Education Association, 1949. 340 p.
2. AMERICAN PSYCHOLOGICAL ASSOCIATION. "Program—1946 Annual Meeting." *American Psychologist* 1: 237-98; July 1946.
3. AMERICAN PSYCHOLOGICAL ASSOCIATION. "Program—1947 Annual Meeting." *American Psychologist* 2: 265-351; August 1947.
4. AMERICAN PSYCHOLOGICAL ASSOCIATION. "Programs of the Division on Evaluation and Measurement." *American Psychologist* 3: 256-62, July 1948; 4: 235-46, July 1949.
5. ARMSTRONG, DAVID T. "Construction of Examinations." *High Points* 30: 41-48; June 1948.
6. BARTLETT, F. C. "The Measurement of Human Skill." *Occupational Psychology* 22: 31-38, 83-91; January, April 1948.
7. BECKLEY, DONALD K., and SMITH, LEO F. "Test Forms for Use in Vocational Education." *Industrial Arts and Vocational Education* 36: 56-59; February 1947.
8. BLATTNER, HELENE. "An Experimental Study of the Testing of Pronunciation." *Speech Monographs* 15: 181-87; No. 2, 1948.
9. BROGDEN, HUBERT E. "Variation in Test Validity with Variation in the Distribution of Item Difficulties, Number of Items, and Degree of Their Intercorrelation." *Psychometrika* 11: 197-214; December 1946.
10. BUROS, OSCAR K., editor. *The Third Mental Measurements Yearbook*. New Brunswick, New Jersey: Rutgers University Press, 1949. 1047 p.

11. CHAUNCEY, HENRY, chairman. *Exploring Individual Differences*. Report of the 1947 Invitational Conference on Testing Problems. American Council on Education Studies, Series I, Number 32. Washington, D. C.: the Council, 1948. 114 p.
12. COUTANT, VICTOR. "Evaluation in Foreign Language Teaching." *Modern Language Journal* 32: 596-99; December 1948.
13. DARLEY, JOHN G., and OTHERS. *The Use of Tests in College*. American Council on Education Studies, Series VI, Number 9. Washington, D. C.: the Council, 1947. 82 p.
14. DAVIS, FREDERICK B. "The Factorial Composition of Two Tests of Comprehension in Reading." *Journal of Educational Psychology* 37: 481-86; November 1946.
15. DAVIS, FREDERICK B. *Item-Analysis Data: Their Computation, Interpretation, and Use in Test Construction*. Harvard Education Papers, No. 2. Cambridge, Massachusetts: Graduate School of Education, Harvard University, 1946. 42 p.
16. DAVIS, FREDERICK B. "Notes on Test Construction: the Reliability of Item-Analysis Data." *Journal of Educational Psychology* 37: 385-90; October 1946.
17. DAYSHAW, LAWRENCE T. "The Measurement of Interest." *Canadian Journal of Psychology* 2: 71-80; June 1948.
18. DONAHUE, WILMA T.; COOMBS, CLYDE H.; and TRAVERS, ROBERT M. W. *The Measurement of Student Adjustment and Achievement*. Ann Arbor: University of Michigan Press, 1949. 256 p.
19. DUNNING, GORDON M. "The Construction and Validation of a Test To Measure Certain Aspects of Scientific Thinking in the Area of First Year College Physics." *Science Education* 33: 221-35; April 1949.
20. FLANAGAN, JOHN C., chairman. *Validity, Norms, and the Verbal Factor*. Proceedings of the 1948 Invitational Conference on Testing Problems. New York: Educational Testing Service, 1949. 117 p.
21. FLETCHER, FRANK M., JR., and HILDRETH, WILLIAM H. "The Evaluation of an English Placement Test." *Educational Research Bulletin* 28: 7-17; January 19, 1949.
22. FRIEDENBERG, EDGAR Z. "The Measurement of Student Insight Into the Structure Underlying the Organization and Viewpoint of the Biological Sciences." *Science Education* 33: 57-64; February 1949.
23. GOOD, CARTER V. "Doctoral Dissertations Under Way in Education." *Phi Delta Kappan* 28: 307-19, March 1947; 29: 305-24, March 1948; 30: 198-220, February 1949.
24. GRIFFIN, ELLA W. "The Development and Evaluation of a Diagnostic Test of Reading Placement for Adults of Limited Reading Ability." *Abstracts of Dissertations and Theses, 1947*. Clark University Bulletin, Vol. 19, No. 182. Worcester, Massachusetts: Clark University, 1948. p. 131-33.
25. GRIGG, AUSTIN E. "A Farm Knowledge Test." *Journal of Applied Psychology* 32: 452-55; October 1948.
26. GROSS, LLEWELLYN. "The Construction and Partial Standardization of a Scale for Measuring Self-Insight." *Journal of Social Psychology* 28: 219-36; November 1948.
27. GUILFORD, JOY P. "Factor Analysis in a Test-Development Program." *Psychological Review* 55: 79-94; March 1948.
28. HARRIS, CHESTER W. "Prediction of the Difficulty Index of Objective Type Spelling Items." *Educational and Psychological Measurement* 7: 319-25; Summer 1947.
29. HEIDER, FRITZ. "On the Construction of a Language Usage Test for the Deaf." *Volta Review* 48: 742-44; November 1946.
30. HORST, PAUL. "Determination of the Optimal Test Length To Maximize the Multiple Correlation." *Psychometrika* 14: 79-88; June 1949.
31. HORST, PAUL. "A Generalized Expression for the Reliability of Measures." *Psychometrika* 14: 21-31; March 1949.
32. HSU, EN-HSI. "Note on and Some Suggested Methods for the Determination of the Validity Coefficient." *Journal of Educational Psychology* 39: 304-307; May 1948.
33. IRWIN, JOHN V. "A Battery of Tests of Speech Hearing." *Speech Monographs* 15: 133-41; No. 2, 1948.
34. JARVIS, JOHN A. "Evaluation of Test Results." *Industrial Arts and Vocational Education* 38: 41-44; February 1949.
35. JOHNSON, ALBERT. "An Index of Item Validity Providing a Correction for Chance Success." *Psychometrika* 12: 51-58; March 1947.

36. JONES, HOWARD L., and SAWYER, MICHAEL O. "New Evaluation Instrument." *Journal of Educational Research* 42: 381-85; January 1949.
37. KENEALLY, KATHERINE G. *Construction and Evaluation of a Diagnostic Test of Study Skills, Grades 4-6*. Studies in Psychology and Psychiatry, Vol. 6, No. 8. Washington, D. C.: Catholic University of America, 1947. 59 p.
38. KINZER, JOHN R., and KINZER, LYDIA G. "The Arithmetic Items in a General Chemistry Placement Examination." *Educational Research Bulletin* 28: 122-27; May 11, 1949.
39. KIRKPATRICK, JAMES J., and CURETON, EDWARD E. "Vocabulary Item Difficulty and Word Frequency." *Journal of Applied Psychology* 33: 347-51; August 1949.
40. LAWSHE, CHARLES N., JR., and MAYER, JAMES S. "Studies in Item Analysis: Effect of Two Methods of Item Validation on Test Reliability." *Journal of Applied Psychology* 31: 271-77; June 1947.
41. LINDQUIST, EVERET F. "The Measurement Book Project of the American Council on Education." *National Projects in Educational Measurement*. American Council on Education Studies, Series I, No. 28. Washington, D. C.: the Council, 1947. p. 1-7.
42. LOEVINGER, JANE. *A Systematic Approach to the Construction and Evaluation of Tests of Ability*. Psychological Monographs, Vol. 61, No. 4. Washington, D. C.: American Psychological Association, 1947. 49 p.
43. LONG, WILLIAM F., and BURR, IRVING W. "Development of a Method for Increasing the Utility of Multiple Correlations by Considering Both Testing Time and Test Validity." *Psychometrika* 14: 137-61; June 1949.
44. MENSCH, IVAN N. "The Effects of Test Abbreviation Upon Responses to the Individual Items." *Journal of Clinical Psychology* 5: 22-36; January 1949.
45. MONTGOMERY, EDITH O. "Construction of Sequence Rank-Test." *Journal of Educational Research* 39: 523-27; March 1946.
46. MOUNT, GEORGE E. "An Efficient Method of Obtaining Counts for Computing the Interrelation of Test Items." *Journal of Applied Psychology* 31: 634-37; December 1947.
47. MURRAY, JOHN E. "An Analysis of Geometric Ability." *Journal of Educational Psychology* 40: 118-24; February 1949.
48. OWENS, WILLIAM A. "An Empirical Study of the Relationship Between Item Validity and Internal Consistency." *Educational and Psychological Measurement* 7: 281-88; Summer 1947.
49. PATTERSON, CHARLES. "On the Problem of the Criterion in Prediction Studies." *Journal of Consulting Psychology* 10: 277-80; September 1946.
50. PEEL, E. A. "Assessment of Interest in Practical Topics." *British Journal of Educational Psychology* 18: 41-47; February 1948.
51. PEEL, E. A. "Item Difficulty as the Measuring Device in Objective Mental Tests." *British Journal of Psychology, Statistical Section* 2: 69-75; July 1949.
52. PEEL, E. A. "Prediction of a Complex Criterion and Battery Reliability." *British Journal of Psychology, Statistical Section* 1: 84-94; July 1948.
53. SEASHORE, HAROLD G., and BENNETT, GEORGE K. "A Test of Stenography: Some Preliminary Results." *Personnel Psychology* 1: 197-209; Summer 1948.
54. SHORES, J. HARLAN. "Some Considerations of Invalidities of General Reading Tests." *Journal of Educational Research* 40: 448-57; February 1947.
55. SIMS, VERNER M. "Questioning Some Assumptions Underlying Current Achievement Testing." *Educational and Psychological Measurement* 8: 565-73; Winter 1948.
56. SMITH, I. MACFARLANE. "Measuring Spatial Ability in School Pupils." *Occupational Psychology* 22: 150-59; July 1948.
57. SPACHE, GEORGE. "The Validity of the Binocular Reading Test." *Journal of Educational Research* 41: 461-66; February 1948.
58. SPITZER, HERBERT F. "Techniques for Evaluating Outcomes of Instruction in Arithmetic." *Elementary School Journal* 49: 21-31; September 1948.
59. SWINEFORD, FRANCES, and HOLZINGER, KARL J. "Selected References on Statistics, the Theory of Test Construction, and Factor Analysis." *School Review* 55: 363-68, June 1947; 56: 361-66, June 1948; 57: 315-20, May-June 1949.
60. THOMSON, GODFREY H. "The Maximum Correlation of Two Weighted Variables." *British Journal of Psychology, Statistical Section* 1: 27-34; October 1947.
61. THORNDIKE, EDWARD L. "The Future of Measurements of Abilities." *British Journal of Educational Psychology* 18: 21-25; February 1948.

62. THURSTONE, LOUIS L. "The Improvement of Examinations." *American Association of University Professors Bulletin* 34: 394-97; June 1948.
63. TROYER, MAURICE E., and ANGELL, GEORGE W. *Manual for the SRA Self-Scorer*. Chicago: Science Research Associates, 1949. 12 p.
64. TURNER, WILLIAM D. "Altruism and Its Measurement in Children." *Journal of Abnormal and Social Psychology* 43: 502-16; October 1948.
65. ULLSVIK, BJARNE R. "An Attempt To Measure Critical Judgment." *School Science and Mathematics* 49: 445-52; June 1949.
66. U. S. ADJUTANT GENERAL'S OFFICE. Staff. *Test Construction Manual—Preliminary Form*. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1947.
67. U. S. DEPARTMENT OF THE ARMY. *How To Make the Picture Test Item*. Washington, D. C.: Superintendent of Documents, U. S. Government Printing Office, 1948. 50 p.
68. VAUGHN, KENNETH W., editor. *National Projects in Educational Measurement*. Report of the 1946 Invitational Conference on Testing Problems. American Council on Education Studies, Series I, No. 28. Washington, D. C.: the Council, 1947. 83 p.
69. VERNON, PHILIP E. "Indices of Item Consistency and Validity." *British Journal of Psychology, Statistical Section* 1: 152-66; November 1948.
70. VERNON, PHILIP E. "The Structure of Practical Abilities." *Occupational Psychology* 23: 81-96; April 1949.
71. VILLARREAL, JESSE J. "A Test of the Aural Comprehension of English for Native Speakers of Spanish." *Speech Monographs* 15: 121-32; No. 2, 1948.
72. WALKER, HELEN M. "Item Selection by Sequential Sampling." *Teachers College Record* 50: 404-409; March 1949.
73. WALL, CLIFFORD N. "Cooperative Testing Program." *American Journal of Physics* 17: 355-56; September 1949.
74. WEAVER, EDWARD K. "Evaluation of Student Achievement in Science." *Science Education* 32: 81-87; March 1948.
75. WEITZMAN, ELLIS, and McNAMARA, WALTER J. *Constructing Classroom Examinations*. Chicago: Science Research Associates, 1949. 140 p.
76. WESMAN, ALEXANDER G. "Effect of Speed on Item-Test Correlation Coefficients." *Educational and Psychological Measurement* 9: 51-57; Spring 1949.
77. WHITE, VERNA. "Measuring Achievement in High School English." *School Review* 55: 474-83; October 1947.

Index to Volume XX, No. 1

Page citations are made to single pages; these are often the beginning of a chapter, section, or running discussion dealing with the topic.

- Achievement tests, 7, 63, 87
- Aptitude tests, 6, 27; art, 34; clerical, 32; differential, 27; mechanical, 31; music, 34; professional, 32; subjectmatter, 30; visual, 34
- Arithmetic tests, 66
- Armed forces, tests in, 8, 28
- Art, and personality, 56
- Art aptitude tests, 34
- Attitudes, measurement of, 46
- Award contests, 76

- Clerical aptitude tests, 32

- Dexterity tests, 31

- Factor analysis, 17, 40, 94

- Geography tests, 66

- Intelligence and achievement, 20; changes in, 22; growth of, 19
- Intelligence tests, 6, 17; applications of, 20; new developments in, 17
- Interests, measurement of, 44
- Interest inventories, 44
- Item analysis, 91

- Language tests, 70

- Mathematics tests, 67
- Measurement (see Tests)
- Mechanical aptitude tests, 31
- Mental ability, growth of, 19
- Music aptitude tests, 34

- Personality, and art, 56; inventories, 38; tests, 38, 51
- Personnel selection, 82
- Prediction of achievement, 21, 82
- Projective technics, 51

- Readability, 70; of interest inventories, 45
- Reading, tests, 65, 67, 70
- Reliability, of personality inventories, 41
- Report cards, 7

- Scholarship contests, 76
- Science tests, 70

- Testing agencies, 9
- Testing programs, 65
- Tests, achievement, 7, 63, 87; administration of, 22; aptitude, 6, 27; arithmetic, 8, 66; in the armed forces, 28; attitude, 46; in college, 68; dexterity, 31; educational, 5; geography, 66; history of, 5; in industry, 81; intelligence, 6, 17; of interests, 44; item analysis, 91; item types, 90; language, 70; mathematics, 67; mechanical aptitude, 31; personality, 7, 38, 51; psychological, 5; publication of, 9; reading, 65, 67, 70; research on, 9; reviews of, 6; science, 70; trends in, 11; use of, 10

- Validity, of achievement tests, 92; of personality inventories, 41
- Visual aptitude tests, 34